INVESTOR IN PEOPLE

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

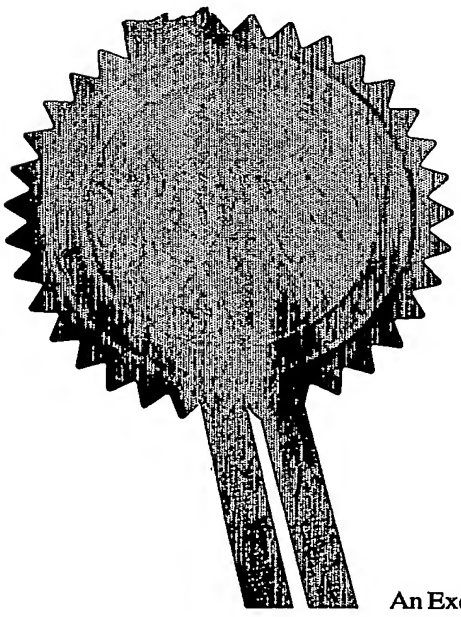
The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed P. Mahoney

Dated 26 June 2003

EP 1/77 03/06757

23.07.03

28JUN02 E729253-1 000524
P01/7700 0.00-0214896.3

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
27 JUN 2002
PERSONS BY HAND

The **Patent Office**

1/77

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP10 8QQ

1.	Your reference	1-32546P1/FMI	
2.	Patent application number *(The Patent Office will fill in this part)*	27 JUN 2002	0214896.3
3.	Full name, address and postcode of the or of each applicant *(underline all surnames)*	NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERLASSUNG FRIEDRICH MIESCHER INSTITUTE FOR BIOMEDICAL RESEARCH MAULBEERSTRASSE 66, CH-4058 BASEL SWITZERLAND	
	Patent ADP number *(if you know it)*	08206583001	
	If the applicant is a corporate body, give the country/state of its incorporation	SWITZERLAND	
4.	Title of invention	Gene for increased somatic recombination	
5.	Name of your agent *(If you have one)*		
	"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	B.A. YORKE & CO. CHARTERED PATENT AGENTS COOMB HOUSE, 7 ST. JOHN'S ROAD ISLEWORTH MIDDLESEX TW7 6NH	
	Patents ADP number *(if you know it)*	1800001	
6.	If you are declaring priority from one ore more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number	Country / Priority application number *(if you know it)*	Date of filing (day/month/year)
7.	If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	Number of earlier application	Date of filing (day/month/year)
8.	Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:* *a) any applicant named in part 3 is not an inventor, or* *b) there is an inventor who is not named as an applicant, or* *c) any named applicant is a corporate body.* *(see note (d))*	Yes	

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	26
Claim(s)	2
Abstract	1
Drawing(s)	

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	ONE
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	Figures 21 pages

11. I/We request the grant of a patent on the basis of this application

Signature B.A. Yorke & Co Date

B.A. Yorke & Co. 27 June 2002

12. Name and daytime telephone number of person to contact in the United Kingdom

Mrs. E. Cheetham
020 8560 5847

Warning

After an application for a patent has been filed. the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the united Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) Once you have filled in the form you must remember to sign and date it.

e) For details of the fee and ways to pay please contact the Patent Office.

DUPLICATE

1-32546P1/FMI

Gene for increased somatic recombination

TECHNICAL FIELD

The present invention relates to DNA that encodes proteins that control somatic recombination, in particular in plants.

BACKGROUND

Cells of all organisms have evolved a series of DNA repair pathways that counteract the deleterious effects of DNA damage and are triggered by intricate signal cascades. Homologous recombination in plants stabilizes the genome by repairing damaged chromosomes simultaneously generating genetic variability through the creation of new genes and new genetic linkages. Repair of DNA damage by recombination is particularly significant for cells under exogenous and endogenous genotoxic stress because of its potential to remove a wide range of DNA lesions. The current understanding of genetic and molecular components underlying meiotic and somatic recombination and DNA repair in plants is limited. To be able to modify or improve DNA repair using gene technology it is necessary to identify key proteins involved in said pathways or cascades.

The precise manipulation of the genome of higher plants still is a major challenge for plant genetic engineering. Some advances have been made recently for the creation of point mutations at predetermined positions by chimeric RNA/DNA oligonucleotides (Beetham et al. 1999, Hohn & Puchta 1999, Zhu et al. 1999, Kipp et al. 2000, Zhu et al. 2000). However, the targeted insertion of longer stretches of DNA sequence at any desired location ("knock-in") or the replacement of predetermined plant genomic sequences by heterologous DNA ("knock-out) via homologous recombination is at present not possible as a routine technique (Mengiste & Paszkowski 1999, Puchta 2002).

Few reports have appeared in the literature that describe successful "gene targeting" in higher plants (Paszkowski et al. 1988, Lee et al. 1990, Offringa et al. 1990, Miao & Lam 1995, Kempin et al. 1997, Hanin et al. 2001), but the reported absolute numbers and relative

frequencies of the desired events were very low. Indeed, the main problem for "gene targeting" experiments is the low frequency of the desired homologous recombination events - 10-3 to 10-5 (Hohn & Puchta 1999, Mengiste & Paszkowski 1999) - relative to illegitimate recombination/integration events.

Various attempts of increasing the low relative frequency of targeted homologous recombination events, by improved selection schemes ("positive-negative selection") or by providing extended regions of sequence homology, were not successful (Thykjaer et al. 1997, Gallego et al. 1999). One promising strategy to facilitate gene targeting in higher plants would be to shift the balance between illegitimate and homologous recombination events towards the latter, by facilitating homologous recombination events in plants by genetic manipulation (Gherbi et al. 2001).

One approach described in the literature is the expression in plants of heterologous proteins known to be involved in homologous recombination. Overproduction of the bacterial resolvase RuvC was shown to increase somatic inter-and intrachromosomal recombination, as well as extrachromosomal recombination (Shalev et al. 1999), but no gene targeting studies were reported yet with this system. Expression of the bacterial RecA protein had similar effects (Reiss et al. 1996, Reiss et al. 1997), but subsequent experiments did not show an increase of gene targeting events (Reiss et al. 2000). So far, it is not clear whether heterologous proteins can successfully interact with the plant recombination machinery to affect the outcome of the recombination events required for gene targeting. In addition, these foreign proteins might have undesired side effects in plants.

An alternative approach is to rely on endogenous plant genes to influence the frequency of homologous recombination events. So far, indirect approaches have been reported to isolate plant genes involved in recombination. The cloning of plant orthologs to recombination and repair genes from other species was reported (Klimyuk & Jones 1997, Doutriaux et al. 1998, Hartung & Puchta 1999, Gallego et al. 2000, Lin et al. 2000), but so far the importance of these genes for recombination in plants has not been evaluated. Functional screens have been carried out to identify plant mutants hypersensitive to genotoxic treatments (Davies et al. 1994, Jenkins et al. 1995, Jiang et al. 1997, Masson et al. 1997, Albinsky et al. 1999, Mengiste et al. 1999). Since recombination is an important mechanism for DNA repair, some of these mutants might be affected in their recombination behavior. This was experimentally

demonstrated for some X-ray hypersensitive Arabidopsis mutants that also showed reduced levels of somatic recombination (Masson & Paszkowski 1997), although the affected gene has not been isolated. Recently, a DNA damage hypersensitive Arabidopsis mutant was isolated from a T-DNA tagged population, the affected gene (MIM) was cloned and shown to encode an SMC (Structural Maintenance of Chromatin) protein. Since the *mim* mutant showed decreased frequencies of somatic recombination, MIM seems be involved in some aspect of somatic recombination (Mengiste et al. 1999). Also in tobacco a hyperrecombinogenic mutant was isolated (Gorbunova et al. 2000). However, the gene affected could not be isolated so far.

Previously, a genetic system was described to study somatic homologous recombination between repeated sequences in whole plants (Swoboda et al. 1994, Puchta et al. 1995a, Puchta et al. 1995b). Briefly, a transgene carrying two non-functional halves of the β-glucuronidase reporter gene sharing a stretch of sequence identity serves as a reporter construct. Homologous recombination between the repeated sequences results in the restoration of a functional reporter gene. Such events were detected by a sensitive histochemical assay, and confirmed by Southern blotting. This assay is destructive, since the staining procedure is lethal, so that direct isolation of mutants is difficult.

Therefore, there is a need in the art to identify genes that increase somatic recombination and this invention meets that need.

RELEVANT LITERATURE

1) Albinsky, D. et al. Plant responses to genotoxic stress are linked to an ABA/salinity signaling pathway. Plant J 17:73-82, 1999.

2) Aravind, L. et al. Conserved domains in DNA repair proteins and evolution of repair systems. Nucl Acids Res 27:1223-1242, 1999.

3) Beetham, P.R. et al. A tool for functional plant genomics: chimeric RNA/DNA oligonucleotides cause in vivo gene-specific mutations. Proc Natl Acad Sci USA 96: 8774-8778, 1999.

4) Cho, S.-G. et al. TIP49b, a regulator of activating transcription factor 2 response to stress and DNA damage. Mol Cell Biol 21:8398-8413, 2001.

5) Clough, S.J. & Bent, A.F. Floral dip: a simplified method for Agrobacterium-mediated

transformation of Arabidopsis thaliana. Plant J 16:735-743, 1998.

6) Davies, C. et al. Isolation of Arabidopsis thaliana mutants hypersensitive to gamma radiation. Mol Gen Genet 243:660-665, 1994.

7) Davis, J.L. et al. A presumptive helicase (MOT1 gene product) affects gene expression and is required for viability in the yeast Saccharomyces cerevisiae. Mol Cell Biol 12:1879-1892, 1992.

8) Dilkes, B.P. & Feldmann, K.A. Cloning genes from T-DNA tagged mutants. Methods Mol Biol 82:339-351, 1998.

9) Doutriaux, M.-P. et al. Isolation and characterisation of the RAD51 and DMC1 homologs from Arabidopsis thaliana. Mol Gen Genet 257:283-291, 1998.

10) Ebbert, R. et al. The product of the SNF2/SWI2 paralogue INO80 of Saccharomyces cerevisiae required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. Mol Microbiol 32:741-751, 1999.

11) Emery, H.S. et al. Sequence of RAD54, a Saccharomyces cerevisiae gene involved in recombination and repair. Gene 104:103-106, 1991.

12) Essers, J. et al. Homologous and non-homologous recombination differentially affect DNA damage repair in mice. EMBO J 19:1703-1710, 2000.

13) Fang, R.X. et al. Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. Plant Cell 1:141-150, 1989.

14) Fridborg, I. et al. The Arabidopsis dwarf mutant shi exhibits reduced gibberellin responses conferred by overexpression of a new putative zinc finger protein. Plant Cell 11:1019-1032, 1999.

15) Gallego, F. et al. AtRAD1, a plant homologue of human and yeast nucleotide excision repair endonucleases, is involved in dark repair of UV damages and recombination. Plant J 21:507-518, 2000.

16) Gallego, M.E. et al. Positive-negative selection and T-DNA stability in Arabidopsis transformation. Plant Mol Biol 39:83-93, 1999.

17) Gamborg, O.L. et al. Nutrient requirement of suspension cultures of soybean root cells. Exp Cell Res 50:151-158, 1968.

18) Gherbi, H. et al. Homologous recombination in planta is stimulated in the absence of Rad50. EMBO Rep 2:287-291, 2001.

19) Gorbalenya, A.E. & Koonin, E.V. Helicases: amino acid sequence comparisons and structure-function relationships. Curr Opin Struct Biol 3:419-429, 1993.

20) Gorbunova, V. et al. A new hyperrecombinogenic mutant of Nicotiana tabacum. Plant J

24:601-611, 2000.

21) Guerineau, F. et al. Sulfonamide resistance gene for plant transformation. Plant Mol Biol 15:127-136, 1990.

22) Hanin, M. et al. Gene targeting in Arabidopsis. Plant J 28:671-677, 2001.

23) Hartung, F. & Puchta, H. Isolation of the complete cDNA of the Mre11 homologue of Arabidopsis (Accession No. AJ243822) indicates conservation of DNA recombination mechanisms between plants and other eucaryotes. (PGR99-132). Plant Physiol 121: 312, 1999.

24) Hayashi, H. et al. Activation of a plant gene by T-DNA tagging: auxinindependent growth in vitro [retracted by Schell J. Science 284:1275, 1999]. Science 258:1350-1353, 1992.

25) Hohn, B. & Puchta, H. Gene therapy in plants. Proc Natl Acad Sci USA 96:8321-8323, 1999.

26) Ikura, T. et al. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell 102:463-473, 2000.

27) Jenkins, M.E. et al. Radiation-sensitive mutants of Arabidopsis thaliana. Genetics 140: 725-732, 1995.

28) Jelesko J.G. et al. Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in Arabidopsis thaliana. Proc Natl Acad Sci USA 96:10302-10307, 1999.

29) Jiang, C.-Z. et al. UV- and gamma-radiation sensitive mutants of Arabidopsis thaliana. Genetics 147:1401-1409, 1997.

30) Jonsson, Z.O. et al. Rvb1p and Rvb2p are essential components of a chromatin remodeling complex that regulates transcription of over 5% of yeast genes. J Biol Chem 276:16279-16288, 2001.

31) Kakimoto, T. CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. Science 274:9825, 1996.

32) Kanemaki, M. et al. TIP49b, a new RuvB-like DNA helicase, is induced in a complex together with another RuvB-like DNA helicase, TIP49a. J Biol Chem 274:22437-22444, 1999.

33) Kardailsky, I. et al. Activation tagging of the floral inducer FT. Science 286:1962-1965, 1999.

34) Kempin, S.A. et al. Targeted disruption in Arabidopsis. Nature 389:802-803, 1997.

35) Kipp, P.B. et al. Gene targeting in plants via site-directed mutagenesis. Methods Mol

Biol 133:213-221, 2000.

36) Klimyuk, V.I. & Jones, J.D. AtDMC1, the Arabidopsis homologue of the yeast DMC1 gene: characterization, transposon-induced allelic variation and meiosis- associated expression. Plant J 11:1-14, 1997.

37) Laurent, B.C. et al. Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. Proc Natl Acad Sci USA 88:2687-2691, 1991.

38) Lee, K.Y. et al. Homologous recombination in plant cells after Agrobacterium-mediated transformation. Plant Cell 2:415-425, 1990.

39) Liu, Z. et al. Repair of UV damage in plants by nucleotide excision repair: Arabidopsis UVH1 DNA repair gene is a homolog of Saccharomyces cerevisiae Rad1. Plant J 21:519-528, 2000.

40) Masson, J.E. et al. Mutants of Arabidopsis thaliana hypersensitive to DNA-damaging treatments. Genetics 146:401-407, 1997.

41) Masson, J.E. & Paszkowski, J. Arabidopsis thaliana mutants altered in homologous recombination. Proc Natl Acad Sci USA 94:11731-11735, 1997.

42) Mathur, J. et al. Gene identification with sequenced T-DNA tags generated by transformation of Arabidopsis cell suspension. Plant J 13:707-716, 1998.

43) Mayerhofer, R. et al. T-DNA integration: a mode of illegitimate recombination in plants. EMBO J 10:697-704, 1991.

44) Mengiste, T. & Paszkowski, J. Prospects for the precise engineering of plant genomes by homologous recombination. Biol Chem 380:749-758, 1999.

45) Mengiste, T. et al. An SMC-like protein is required for efficient homologous recombination in Arabidopsis. EMBO J 18:4505-4512, 1999.

46) Miao, Z.H. & Lam, E. Targeted disruption of the TGA3 locus in Arabidopsis thaliana. Plant J 7:359-365, 1995.

47) Michelet, B. & Chua, N.-H. Improvement of Arabidopsis mutant screens based on luciferase imaging in planta. Plant Mol Biol Rep 14:320-329, 1996.

48) Millar, A.J. Circadian clock mutants in Arabidopsis identified by luciferase imaging. Science 267:1161-1163, 1995a.

49) Millar, A.J. et al. Firefly luciferase as a reporter of regulated gene expression in higher plants. Plant Mol Biol Rep 10:324-327, 1992.

50) Millar, A.J. et al. The regulation of circadian period by phototransduction pathways in Arabidopsis. Science 267:1163-1166, 1995b.

51) Muchardt, C. & Yaniv, M. ATP-dependent chromatin remodelling: SWI/SNF and Co.

are on the job. J Mol Biol 293:187-98, 1999.

52) Muris, D.F. et al. Isolation of the Schizosaccharomyces pombe RAD54 homologue, rhp54+, a gene involved in the repair of radiation damage and replication fidelity. J Cell Sci 109:73-81, 1996.

53) Nacry, P. et al. Major chromosomal rearrangements induced by T-DNA transformation in Arabidopsis. Genetics 149:641-650, 1998.

54) Offringa, R. et al. Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation. EMBO J 9:3077-3084, 1990.

55) Paszkowski, J. et al. Gene targeting in plants. EMBO J 7:4021-4026, 1988.

56) Pruitt, R.E. & Meyerowitz, E.M. Characterization of the genome of Arabidopsis thaliana. J Mol Biol 187:169-183, 1986.

57) Puchta, H. et al. Somatic intrachromosomal homologous recombination events in populations of plant siblings. Plant Mol Biol 28:281-292, 1995a.

58) Puchta, H. et al. Induction of intrachromosomal homologous recombination in whole plants. Plant J 7:203-210, 1995b.

59) Puchta, H. Gene replacement by homologous recombination in plants. Plant Mol Biol 48:173-182, 2002.

60) Reiss, B. et al. RecA protein stimulates homologous recombination in plants. Proc Natl Acad Sci USA 93:3094-3098, 1996.

61) Reiss, B. et al. Targeting of a functional Escherichia coli RecA protein to the nucleus of plant cells. Mol Gen Genet 253:695-702, 1997.

62) Reiss, B. et al. RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by agrobacterium. Proc Natl Acad Sci USA 97:3358-3363, 2000.

63) Richmond, E. & Peterson, C.L. Functional analysis of the DNA-stimulated ATPase domain of yeast SWI2/SNF2. Nucl Acids Res 24:3685-92, 1996.

64) Schaffer, R. et al. The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. Cell 93:1219-29, 1998.

65) Shalev, G. et al. Stimulation of homologous recombination in plants by expression of the bacterial resolvase ruvC. Proc Natl Acad Sci USA 96:7398-7402, 1999.

66) Shen, X. et al. A chromatin remodeling complex involved in transcription and DNA processing. Nature 406:541-544, 2000.

67) Simon, M. et al. The 3' to 5' exonuclease activity located in the DNA polymerase δ subunit of Saccharomyces cerevisiae is required for accurate replication. EMBO J 10:

2165-2170, 1991.

68) Sitney, K.C. et al. DNA polymerase III, a second essential DNA polymerase, is encoded by the S. cerevisiae CDC2 gene. Cell 56:599-605, 1989.

69) Sugino, A. Yeast DNA polymerases and their role at the replication fork. TIBS 20:319-323, 1995.

70) Swoboda, P. et al. Intrachromosomal homologous recombination in whole plants. EMBO J 13:484-489, 1994.

71) Thykjaer, T. et al. Gene targeting approaches using positive-negative selection and large flanking regions. Plant Mol Biol 35:523-530, 1997.

72) Torres-Ramos, C.A. et al. Requirement of Yeast DNA polymerase δ in post-replicational repair of UV-damaged DNA. J Biol Chem 272:25445-25448, 1997.

73) Travers, A. An engine for nucleosome remodeling. Cell 96:311-314, 1999.

74) Troelstra, C. et al. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell 71:939-53, 1992.

75) van Gool, A.J. et al. RAD26, the functional S. cerevisiae homolog of the Cockayne syndrome B gene ERCC6. EMBO J 13:5361-5369, 1994.

76) Walden, R. et al. Activation tagging: a means of isolating genes implicated as playing a role in plant growth and development. Plant Mol Biol 26:1521-1528, 1994.

77) Weigel, D. et al. Activation tagging in Arabidopsis. Plant Physiol 122:1003-1013, 2000.

78) Wilson, K. et al. A Dissociation insertion causes a semidominant mutation that increases expression of TINY, an Arabidopsis gene related to APETALA2. Plant Cell 8: 659-671, 1996.

79) Wood, M. et al. An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. Mol Cell 5:321-330, 2000.

80) Zhu, T. et al. Targeted manipulation of maize genes in vivo using chimeric RNA/DNA oligonucleotides. Proc Natl Acad Sci USA 96:8768-8773, 1999.

81) Zhu, T. et al. Engineering herbicide-resistant maize using chimeric RNA/DNA oligonucleotides. Nat Biotechnol 18:555-558,2000.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts sequences related to mbm17.5 A. predicted cDNA of mbm17.5 B. predicted protein sequence of MBM17.5 C. full length cDNA of mbm17.5 D. protein sequence of MBM17.5 E. Over-expressed transcript of mbm17.5 in mutant hw17

Figure 2 depicts sequences related to mbm17.6 A. predicted cDNA of mbm17.6 (DNA polymerase III) B. predicted protein sequence of MBM17.6 (DNA polymerase III)

Figure 3 depicts the partial sequence of Osmbm17.5>EST clone RICS1367A, Oryza sativa homolog of mbm17.5

Figure 4 depicts the partial sequence of zmmbm17.5>EST clone 603011H11, Zea mays homolog of mbm17.5

Figure 5 depicts AtIno80 sequence and related sequences A. AtIno80 coding sequence B. AtIno80 derived protein sequence C. Alignment of AtIno 80 sequence and public sequence, At3g57300, showing splicing difference ("Query" refers to AtIno80 sequence; "Sbjct" to public database sequence, gi|18410689|ref|NM_115590.1| (AGI:At3g57300)

Figure 6 depicts the nucleotide sequences of AtRvb1 (At5g22330)

Figure 7 depicts the nucleotide sequences of AtRvb21 (At5g67630)

Figure.8 depicts the nucleotide sequences of AtRvb22 (At3g49830)

Figure 9 depicts the nucleotide sequences of At3g57290

Figure 10 depicts the alignment of protein sequences from MBM17.5, zmMBM17.5 and osMBM17.5, helicase motifs are marked as squares

SUMMARY OF THE INVENTION

The present invention provides an isolated nucleic acid, in particular DNA, comprising a sequence having 98.5% or more identity with the sequences depicted in Figure 1C, Figure 1E or Figure 5A. Also provided are vectors and host cells comprising the nucleic acids of the invention, as well as polypeptides encoded by the nucleic acids.

In a further aspect of the invention, a method for inducing homologous recombination in a cell is provided, comprising modulating the expression or properties of one or more gene

products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21 (At5g67630), Rvb22 (At3g49830) and At3g57290, their homologues, fragments or derivatives. In one embodiment, modulation is achieved by increasing expression of the gene product, such as by introducing a nucleic acid encoding the gene product into the cell operably linked to a promoter; and allowing transcription and translation of the gene in an amount sufficient to affect homologous recombination in said cell.

The method can be used to increase somatic homologous recombination and/or meiotic homologous recombination. The promoter can be an inducible promoter, a tissue-specific promoter, a constitutive promoter or a meiosis-specific promoter, depending on the desired effect.

Also provided is a method of increasing gene targetting to a desired locus in a host cell comprising introducing a desired gene into a host cell, modulating the expression or properties of one or more gene products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21 (At5g67630), Rvb22 (At3g49830) and At3g57290, or functional fragments, derivatives and homologues thereof in the host cell, and detecting integration of the desired gene at a selected locus in the genome of the host cell.

DETAILED DESCRIPTION OF THE INVENTION

The present inventors have used a direct screening approach to identify mutants of *Arabidopsis thaliana* showing increased frequencies of somatic recombination, by visualizing recombination events in living plants from a mutagenized population and directly isolating plants with the desired phenotype. The description below describes a genetic screen and two Arabidopsis mutants *hw17* and *sm22* derived from it, and the associated plant genes responsible for the altered recombination phenotype.

Existing technologies for gene targeting in plants are very inefficient. The modulation of the expression or properties of one or more gene products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330) and Rvb2(1 and 2; also referred to herein as Rvb21 or At5g67630, and Rvb22 or At3g49830, respectively and At3g57290, increases the efficiency of gene targeting events and facilitates

the routine manipulation of the genome of higher plants by homologous recombination. For the purposes of this disclosure, to avoid repetition, reference to the above group of gene products is meant to include reference to each gene individually, i.e., the modulation of the expression or properties of MBM17.5, the modulation of the expression or properties of MBM17.6, and so on.

An *in vivo* screen for *Arabidopsis* mutants has been devised to allow direct detection of mutants with increased recombination. As a result of the screen, and mutant plants with a more than 10-fold increased or altered frequency of somatic recombination events are provided, as well as the plant genes, MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21 (At5g67630), Rvb22 (At3g49830) and At3g57290 affected in these mutant plants, and orthologs from other plant species. The screen allows the identification of mutant plants, and plant genes with a strong effect on recombination having little or no undesired side effects on the plant. An increase in homologous recombination frequency is useful to achieve an increased efficiency of gene targeting in plants.

Within the context of the present invention reference to a gene is to be understood as reference to a DNA coding sequence associated with regulatory sequences, which allow transcription of the coding sequence into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Examples of regulatory sequences are promoter sequences, 5' and 3' untranslated sequences, introns, and termination sequences.

A promoter is understood to be a DNA sequence initiating transcription of an associated DNA sequence, and may also include elements that act as regulators of gene expression such as activators, enhancers, or repressors.

Expression of a gene refers to its transcription into RNA or its transcription and subsequent translation into protein within a living cell. In the case of antisense constructs expression refers to the transcription of the antisense DNA only.

The term transformation of cells designates the introduction of nucleic acid into a host cell, particularly the stable integration of a DNA molecule into the genome of said cell.
Any part or piece of a specific nucleotide or amino acid sequence is referred to as a

component sequence or fragment.

In one aspect of the invention, nucleic acids and polypeptides are provided that can modulate homologous recombination. A nucleic acid according to the present invention comprises a sequence having 98.5%, 99%, 99.5% or more identity with the sequences depicted in Figure 1C, Figure 1E or Figure 5A. The DNA sequence in Figure 1A is 99.8% identical to Figure 1C, due to the different splicing. The nucleic acid can be DNA or RNA, such as, mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Also provided is a vector comprising the nucleic acid of the invention, as well as host cells comprising the vector or nucleic acid of the invention. Suitable vectors and host cells are described in more detail below. Also provided are polypeptides encoded by the nucleic acids of the invention.

In a further aspect of the invention, methods for increasing homologous recombination are provided by modulating the expression or properties of one or more gene products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21 (At5g67630), Rvb22 (At3g49830) and At3g57290. In order to increase homologous recombination several methods are useful depending on the gene and the gene targeting technique employed. Typically, modulation will mean increasing the activity of the gene product, which can easily be achieved by methods known in the art.

In one embodiment, the desired gene is overexpressed in a host cell in an amount sufficient to increase homologous recombination in the host cell. By "overexpression", it is meant increasing the amount of desired gene product in a host cell, compared to untreated cells. A simple way to achieve overexpression is to produce transgenic host cells, in particular transgenic plants, carrying a construct (vector) that ectopically overexpresses the sequence of interest under the control of a suitable promoter, such as the 35S CaMV, MAS (mannopine synthase) or ubiquitin promoter.

In another embodiment, an inducible promoter is used to allow an increase in homologous recombination frequency at the time and place needed, for example, for gene targeting.

Alternatively, the construct increasing recombination can be provided at the same time as the targeting construct by co-transformation, the effect is then achieved by the transient expression of the construct containing the said genes.

It will be apparent to one of ordinary skill in the art that functional fragments, homologues or derivatives of the desired gene can be used. Dynamic programming algorithms yield different kinds of alignments. In general there exist two approaches towards sequence alignment. Algorithms as proposed by Needleman & Wunsch and by Sellers align the entire length of two sequences providing a global alignment of the sequences. The Smith-Waterman algorithm on the other hand yields local alignments. A local alignment aligns the pair of regions within the sequences that are most similar given the choice of scoring matrix and gap penalties. This allows a database search to focus on the most highly conserved regions of the sequences. It also allows similar domains within sequences to be identified. To speed up alignments using the Smith-Waterman algorithm both BLAST (Basic Local Alignment Search Tool) and FASTA place additional restrictions on the alignments.

Within the context of the present invention alignments are conveniently performed using BLAST, a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. Version BLAST 2.0 (Gapped BLAST) of this search tool has been made publicly available on the internet (currently http://www.ncbi.nlm.nih.gov/BLAST/). It uses a heuristic algorithm which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences which share only isolated regions. The scores assigned in a BLAST search have a well-defined statistical interpretation. Particularly useful within the scope of the present invention are the blastp program allowing for the introduction of gaps in the local sequence alignments and the PSI-BLAST program, both programs comparing an amino acid query sequence against a protein sequence database, as well as a blastp variant program allowing local alignment of two sequences only. Said programs are preferably run with optional parameters set to the default values.

For example, GenBank database annotation of mbm17.5 predicted a gene with similarities to Rad26 nucleotide excision repair proteins. Comparision of the predicted protein-coding segments against the GenPept/SwissProt protein database using the BLASTP program revealed many similar protein sequences of known function of the SWI2/SNF2 helicase/ATPase protein family. A similarity search of the protein database revealed that the central region of this predicted protein of 1043 amino acids has significant similarity to a number of proteins involved in DNA binding, repair, recombination, and chromatin

remodeling. In particular, the human protein ERCC6 (Troelstra et al. 1992), involved in Cockayne's syndrome, and its S. cerevisiae homologue RAD26 (van Gool et al. 1994) are important for the repair of active genes, the proteins RAD54 and rph54, from S. cerevisiae and S. pombe (Emery et al. 1991, Muris et al. 1996) and their mammalian homologues (Essers et al. 2000) are involved in DNA recombination and repair, and the yeast proteins MOT1 (Davis et al. 1992) and SNF2 (Laurent et al. 1991, Richmond & Peterson 1996) are known to affect the expression of numerous genes, most likely by ATP-dependent chromatin remodeling. All these proteins share an extended protein sequence motif with the predicted product of the MBM17.5 coding sequence, the so-called helicase/ATPase domain of the SWI2/SNF2 protein family (Gorbalenya & Koonin 1993, Aravind et al. 1999, Muchardt & Yaniv 1999, Travers 1999) and may be useful in increasing homologous recombination frequency.

Sequence alignments using BLAST can also take into account whether the substitution of one amino acid for another is likely to conserve the physical and chemical properties necessary to maintain the structure and function of the protein or is more likely to disrupt essential structural and functional features of a protein. Such sequence similarity is quantified in terms of a percentage of "positive" amino acids, as compared to the percentage of identical amino acids and can help assigning a protein to the correct protein family in border-line cases.

Specific examples of DNA and encoded proteins according to the present invention are described in Figures 1, 2, 3, 4, 5, 6, 7, 8 and 9. Typically, functional fragments or derivatives are characterized by an amino acid sequence comprising a component sequence of at least 150 amino acid residues having 40% or more identity with an aligned component sequence of the one or more of the polypeptides encoded by the DNA of Figures 1 to 9. Preferably the amino acid sequence identity is higher than 50% or even higher than 55%.

DNA encoding proteins according to the present invention can be isolated from monocotyledonous and dicotyledonous plants. Preferred sources are corn, sugarbeet, sunflower, winter oilseed rape, soybean, cotton, wheat, rice, potato, broccoli, cauliflower, cabbage, cucumber, sweet corn, daikon, garden beans, lettuce, melon, pepper, squash, tomato, or watermelon. However, they can also be isolated from mammalian sources such as mouse or human tissues. The following general method, can be used, which the person

skilled in the art knows to adapt to the specific task. A single stranded fragment of the desired gene consisting of at least 15, preferably 20 to 30 or even more than 100 consecutive nucleotides is used as a probe to screen a DNA library for clones hybridizing to said fragment. The factors to be observed for hybridization are described in Sambrook et al, Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, chapters 9.47-9.57 and 11.45-11.49, 1989. Hybridizing clones are sequenced and DNA of clones comprising a complete coding region encoding a protein characterized by an amino acid sequence comprising a component sequence of at least 150 amino acid residues having 40% or more sequence identity to the protein sequence encoded by the desired gene is purified. Said DNA can then be further processed by a number of routine recombinant DNA techniques such as restriction enzyme digestion, ligation, or polymerase chain reaction analysis. The disclosure of the nucleotide sequences in Figs 1-9 enables a person skilled in the art to design oligonucleotides for polymerase chain reactions which attempt to amplify DNA fragments from templates comprising a sequence of nucleotides characterized by any continuous sequence of 15 and preferably 20 to 30 or more basepairs of the desired gene.

Suitable vectors for practicing the methods of the invention are well known in the art. Similalry, host cells can be derived from monocotyledonous or dicotyledonous plants. Preferred sources are corn, sugarbeet, sunflower, winter oilseed rape, soybean, cotton, wheat, rice, potato, broccoli, cauliflower, cabbage, cucumber, sweet corn, daikon, garden beans, lettuce, melon, pepper, squash, tomato, or watermelon. However, host cells can also be isolated from other sources, including mammalian sources such as mouse or human cells, in particular stem cells. It is preferred that mammalian homologues are used in mammalian cells.

The methods for increasing homologous recombination are useful to obtain gene targeting so that a gene of interest is introduce into the genome at a desired locus, instead of randomly. For some hosts, in particular crop plants, the gene is preferably expressed in a selected tissue where expression is needed. This is easily achieved by the use of tissue specific promoter. Thus, the present invention provides a method for increasing somatic homologous recombination and increasing gene targetting by modulating the expression or properties of one or more gene products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21

(At5g67630), Rvb22 (At3g49830) and At3g57290, and fragments, derivatives and homologues thereof, essentially as described above.

The methods are also useful to improve meiotic recombination, thereby facilitating breeding of species, in which genes encoding a particular phenotype are transferred between plants. Crossing in an interesting trait from another variety or species into a given variety by conventional breeding is a very time and labour-intensive process. Several generations of back-crosses have to be carried out to eliminate the undesired genetic material of the donor species, while maintaining the desired phenotype or trait. Using the methods described above for increasing homologous recombination, meiotic recombination frequencies can be increased, preferably by expressing the desired gene under the control of a meiosis-specific promoter or inducible promoter, the breeding process is speeded up. Thus, the present invention provides a method for increasing meiotic recombination by modulating the expression or properties of one or more gene products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21 (At5g67630), Rvb22 (At3g49830) and At3g57290, and fragments, derivatives and homologues thereof, essentially as described above.

The Examples below are provided for illulstrative purposes and are in no way intended to be limiting to the invention.

EXAMPLES:

Example 1: Identification of At5g63950 (MBM17.5) as gene effective in increasing homologous gene recombination in the mutant hw17.

We have used for our screening a newly constructed a transgenic Arabidopsis thaliana line that carries a recombination reporter construct based on the firefly luciferase gene. The structure of the reporter construct - two segments of the luciferase gene arranged as inverted repeats - is comparable to that of the previously described beta-glucuronidase reporter (Swoboda et al. 1994, Puchta et al. 1995a, Puchta et al. 1995b), but offers the advantage that recombination events can be detected in living plants. Luciferase activity in cells in which recombination has restored an intact luciferase gene can be detected by light emission after application of the substrate D-luciferin using a high-sensitivity CCD camera (Millar et al. 1992, Millar et al. 1995a, Millar et al. 1995b, Michelet & Chua 1996).

To induce hyperrecombination mutations in the luciferase recombination reporter line, we used T-DNA activation tagging with a mutagenic construct (pAC102). "Activation tagging" refers to the transcriptional activation of endogenous plant genes by random integration of a construct that carries promoter or enhancer sequences. One published approach for "activation tagging" is the introduction, via Agrobacterium-mediated gene transfer, of a T-DNA carrying several copies of the cauliflower mosaic virus (CaMV) 35S enhancer (Fang et al. 1989), which can activate the expression of heterologous genes over a distance (Hayashi et al. 1992, Walden et al. 1994, Kakimoto 1996, Kardailsky et al. 1999, Weigel et al. 2000). Another published approach is the introduction of a complete, outward-pointing CaMV 35S promoter on a transposable Ds element (Wilson et al. 1996, Schaffer et al. 1998, Fridborg et al. 1999). The construct "pAC102" used for our experiments is a combination of these previously described elements: it is a binary vector carrying a T-DNA that can be transferred to plants that contains a complete, outward-pointing copy of the CaMV 35S promoter/enhancer close to the right T-DNA border. Thus, this construct combines the ease of application of T-DNA gene transfer with the genetic ability of a complete promoter, avoiding some of the drawbacks of enhancer-only constructs (Weigel et al. 2000).

In principle, the activation tagging construct can cause several kinds of mutations after integration in the plant genome: gene disruption by insertion within a coding sequence, activation of plant gene expression by action of the CaMV 35S enhancer, direct expression of a plant gene from the CaMV 35S promoter on the T-DNA, or down-regulation of expression by antisense RNA production driven from the CaMV 35S promoter. The pAC102 T-DNA carries in addition to the 35S promoter a complete copy of the pUC cloning vector to facilitate gene cloning by plasmid rescue (Dilkes & Feldmann 1998), and a sulfonamide resistance marker (Guerineau et al. 1990, Reiss et al. 1996) for selection of transgenic plants.

We transformed 13.000 three-week old Arabidopsis ecotype Columbia plants from the luciferase recombination reporter line with the activation-tagging T-DNA construct "pAC102" by Agrobacterium-mediated gene-transfer, using the established "floral dip" procedure (Clough & Bent 1998) with a modified infiltration buffer, in which the Silwet L-77 detergent was replaced by 0.05% Extravon® (Ciba). Seeds from the infiltrated plants were harvested three weeks after infiltration. Transgenic progeny carrying the pAC102 activation tagging T-

DNA were selected by sowing seeds on perlite substrate drenched with Gamborg B5 mineral medium (Gamborg et al. 1968) containing 10 mg/l sulfadiazine (Sigma), and transferring surviving individuals after 10 days to soil. About 20,000 sulfonamide-resistant plants were isolated; they represent independent transformants with the pAC102 T-DNA activation tagging construct integrated at different random positions in the Arabidopsis genome.

When individual plants had grown to the 10-leaf stage, they were assayed for luciferase activity to detect somatic recombination events. Batches of 25 plants were sprayed with the substrate D-luciferin and pictures (typically two) were taken with a "Astrocam" (Gloor Instruments, Uster) by integrating photons over 15 min. Background noise and cosmic radiation was filtered out by correlating both images using the minimum function. Plants showing an increased number of sectors with luciferase activity relative to the average of the population were observed with a frequency of about 1 in 500 plants.

As a result of the screen, one plant line, termed "hw17", showed a more than 10-fold increase in number of luciferase sectors. The original transformed plant "hw17" was grown to maturity to obtain seeds. Progeny plants also showed an increase in number of luciferase-expressing sectors, suggesting that this plant line carries a heritable mutation resulting in increased somatic recombination frequencies. To characterize the T-DNA integration pattern in this plant line by Southern blotting, callus was induced from leaves of the original transformant, and genomic DNA was prepared once sufficient plant material was produced. DNA was digested with HindIII, transferred to nylon membranes after electrophoretic separation, and probed using a DIG-labeled pUC-bla (pUC beta lactamase) PCR product to detect genomic fragments carrying the right end of the pAC102 activation tagging T-DNA. HindIII cuts twice within the pAC102 T-DNA, and the pUC-bla segment detected by the probe lies between the T-DNA right border and one of these recognition sites. Therefore, each independent integration site is detected as a HindIII fragment on Southern blots consisting of the right end of the pAC102 T-DNA, including the pUC vector sequences and the CaMV 35S promoter, and a variable length of plant DNA extending from the right end integration site of the pAC102 T-DNA to the nearest HindIII restriction site in the plant genome.

Two bands of about 5 kb and 10 kb were detected, suggesting two independent T-DNA insertion events. To isolate the plant genomic sequences adjacent to the pAC102 T-DNA

integration, we used the technique of plasmid rescue cloning (Dilkes & Feldmann 1998, Mathur et al. 1998). Briefly, we digested plant genomic DNA with HinDIII, circularized the resulting fragments by ligation at low DNA concentration, and transformed the ligation mixture into competent *E. coli* TOP10 cells (commercially available from INVITROGEN) by electroporation. Since the HindIII fragments containing the fusion joint between plant DNA and the right end of the activation tagging construct carry a plasmid origin and the ampicillin resistance gene (bla) contributed by pAC102, circularization of such fragments will result in a functional bacterial plasmid and confer ampicillin resistance to the *E. coli* cells.

Several colonies were obtained after plating the transformed bacteria on selection medium containing ampicillin. Plasmid DNA of these transformants was prepared and characterized by restriction analysis. The plasmids fell into two classes after re-digestion with HindIII: one class contained a HindIII fragment of approximately 5 kb, the other one of approximately 10 kb; corresponding to the size of the T-DNA integration fragments in the genome of the hw17 plant detected by Southern blotting described above.

To determine the nature of the plant sequences joined to the right end of the T-DNA, the plant DNA insert from these rescued plasmids was sequenced from both sides, using one custom sequencing primer complementary to the T-DNA right end reading towards the plant DNA, and the standard M13 reverse sequencing primer, reading from the pAC102 vector sequences into the plant DNA insert from the other end. The obtained DNA sequences were compared to the GenBank nucleotide database using the BLASTN search program.

The insert of one plasmid, pJL604.2, corresponding in size to the 10 kb band detected on the Southern blots, was highly similar to several Arabidopsis genomic ribosomal DNA gene repeat sequences. This suggests that one of the two pAC102 copies detected in the genome of the hyperrecombination mutant plant "hw17" is located within rDNA repeats. There are about 570 highly expressed copies of these sequences distributed throughout the Arabidopsis genome (Pruitt & Meyerowitz 1986), therefore we consider it very unlikely that changes of expression or mutation of one of them caused by an insertion of the activation tagging construct would cause a hyperrecombination phenotype.

The insert of a second plasmid, pJL604.1, was identical to part of a 52717 bp P1 clone (MBM17) derived from chromosome 5 of Arabidopsis thaliana (GenBank Accession number

AB019227; submitted on 29-OCT-1998 to the DDBJ/EMBL/GenBank databases by Yasukazu Nakamura, Kazusa DNA Research Institute). The sequence contained in the circularized rescued plasmid pL604.1 extends from nucleotide 20310 of MBM17, that is joined to the pAC102 right end with the 35S promoter, to a HindIII site at position 18503 in MBM17, that is joined to an internal HindIII site within pAC102. To confirm that the T-DNA - plant DNA junction found on plasmid pJL604.1 really is derived from the genome of the hyperrecombination mutant "hw17", we performed a PCR reaction with one primer annealing within the Arabidopsis genomic insert and one annealing close to the right border of the pAC102 T-DNA. Using pJL604.1 plasmid DNA or "hw17" plant genomic DNA, we observed a PCR product of identical size, confirming that pJL604.1 carries the authentic pAC102 - plant DNA fusion joint.

In the mutant plant "hw17", the right end of the pAC102 activation tagging T-DNA is fused to nucleotide 20310 of the plant genomic sequence, in such a way that the 35S promoter is pointing towards the beginning of the genomic clone MBM17. Further characterization of the genetic locus revealed complex rearrangments of DNA upon integration of the T-DNA. In particular, genomic Arabidopsis DNA was found inserted into the coding region of the predicted gene mbm17.6.

It has been reported that T-DNA insertions are often accompanied by small deletions or rearrangement of DNA sequences in the vicinity of the T-DNA (Mayerhofer et al. 1991; Nacry et al. 1998). Also, the enhancer present in the CaMV 35S promoter could affect the expression of genes over a distance and might act on several genes surrounding the integration site, although so far enhancer action observed in Arabidopsis plants in vivo in activation tagging experiments did not affect sequences further than 3.6 kb away (Weigel et al. 2000).

GenBank database annotation of mbm17.5 predicted a gene with similarities to Rad26 nucleotide excision repair proteins (Figure 1). Comparision of the predicted protein-coding segments against the GenPept/SwissProt protein database using the BLASTP program revealed many similar protein sequences of known function of the SWI2/SNF2 helicase/ATPase protein family. A similarity search of the protein database revealed that the central region of this predicted protein of 1043 amino acids has significant similarity to a number of proteins involved in DNA binding, repair, recombination, and chromatin

remodeling. In particular, the human protein ERCC6 (Troelstra et al. 1992), involved in Cockayne's syndrome, and its S. cerevisiae homologue RAD26 (van Gool et al. 1994) are important for the repair of active genes; the proteins RAD54 and rph54, from S. cerevisiae and S. pombe (Emery et al. 1991, Muris et al. 1996) and their mammalian homologues (Essers et al. 2000) are involved in DNA recombination and repair, and the yeast proteins MOT1 (Davis et al. 1992) and SNF2 (Laurent et al. 1991, Richmond & Peterson 1996) are known to affect the expression of numerous genes, most likely by ATP-dependent chromatin remodeling. All these proteins share an extended protein sequence motif with the predicted product of the MBM17.5 coding sequence, the so-called helicase/ATPase domain of the SWI2/SNF2 protein family (Gorbalenya & Koonin 1993, Aravind et al. 1999, Muchardt & Yaniv 1999, Travers 1999).

We consider it most likely that the hyperrecombination phenotype detected in mutant line "hw17" is caused by insertion of the activation T-DNA into the predicted coding sequence MBM17.5. Since the recombination phenotype was observed in primary transformants, it is most likely dominant. An insertion of the pAC102 T-DNA at the observed position might cause a phenotype by disrupting the coding sequence of MBM17.5 and/or by the overexpression of a C-terminal fragment of this coding sequence that might have some activity by itself, or might interfere with the function of the intact MBM17.5 gene product. Using Northern Blot analysis and RT-PCR technology we have shown that the activation tag of the pAC102 is active in the mutant hw17, giving rise to a very abundant transcript (Figure 1) with a 705 bp open reading frame, homologous to the last 235 amino acids of the MBM17.5 protein. Although not wishing to be bound by theory, this truncated polypeptide may cause of the hyperrecombination phenotype of the mutant hw17 by sequestering out the functional, complete gene product.

Because of its strong similarity with other proteins known to be involved in DNA repair, chromatin structure and recombination, we consider that the MBM17.5 predicted coding sequence is the target for the mutation in the hyperrecombination mutant plant "hw17". The DNA sequence is a useful tool to manipulate somatic recombination in Arabidopsis. For example, over-expression of the truncated C-terminus of MBM17.5 is dominant, therefore allowing recombination frequency to be manipulated in selected cells by the use of tissue-specific promoters and/or transiently by use of inducible promoters.

The sequence of the cloned, full length cDNA of the mbm17.5 gene (Figure 1) encodes a protein of 1090 amino acids. There are two differences between the cloned and predicted protein sequence, due to the use of different splice sites *in vivo* than in the predicted transcript. Using sequence alignment algorithms we found that the highest similarity to known proteins is restricted to the central part of mbm17.5 (aa 370- aa 900), containing the seven conserved helicase/ATPase motifs of the SWI2/SNF2 helicase family. The amino- and carboxy-termini of the predicted protein MBM17 seem to be less strongly conserved. Orthologs of MBM17.5 in other plant species have been identified (see example 3 and 4).

The link between the hyperrecombination phenotype and the T-DNA insertion in MBM17.5 has been confirmed by segregation analysis of progeny up to the T6 generation. Analyses of plants over-expressing the cDNA, parts of it or anti-sense sequences can be used to demonstrate that the alteration of somatic homologous recombination frequency is due to mbm17.

Example 2: Identification of At5g63960 (mbm17.6) as gene causing hyperrecombination in mutant hw17

Due to the complex rearrangements upon integration of the mutagenizing T-DNA, other genes in the region of BAC clone mbm17 could be affected. While cloning of the T-DNA left border- genomic DNA cloning we found an insertion of Arabidopsis genomic DNA, located on TAC clone K19M22, into the coding region of the neighboring gene mbm17.6. The DNA is integrated 6 nucleotides down-stream of the putative start codon, probably abolishing the expression of a functional gene product of mbm17.6.

GeneBank annotation of mbm17.6 predicted a gene (Figure 2A) homologous to the DNA polymerase III, catalytic subunit of *S. cerevisiae* (Sitney et al. 1989). DNA polymerase III was shown to be involved in the accurate DNA replication (Simon et al. 1991; for review: Sugino 1995) and in post-replicational repair of damaged DNA (Torres-Ramos et al. 1997). DNA replication and repair pathways are dependent on DNA polymerases, so the hyperrecombination phenotype of hw17 could be caused by the presence of less or non-functional DNA polymerase III protein.

Example 3: The rice homolog of mbm17.5 can be used to increasing homologous gene recombination

Targeted genetic modification of the model plant Arabidopsis might become an important tool for academic research but the need for targeted gene placement is much higher in crop plants. Using t-Blast algorithm to seek plant EST database, we found rice (RICS1367A; MAFF DNA bank, Japan) EST clones. We sequenced the rice EST clone RICS1367A and found an open reading frame having high homology to the MBM17.5 protein sequence covering not only the conserved helicase/ATPase motifs (Figs. 3, 10) but extending to the C-terminus of mbm17.5. The rice homolog of mbm17.5 can be used for increasing the efficiency of targeted modification of rice plants, following the strategies described earlier.

Example 4: The maize homolog of mbm17. 5 can be used to increase homologous gene recombination

Using t-Blast algorithm to seek plant EST database, we found a maize (603011H11; Stanford University, USA) EST clone that is an ortholog of the Arabidopsis gene mbm17.5 (Figs. 4, 10). The maize homolog gene of mbm17.5 can be used to increase the efficiency of targeted modification of maize, following the strategies described above.

Example 5: *sm22 mutant.* Determination of sm22 transcript (helicase/ATPase) as an agent that improves homologous recombination

From the same screen as described in Example 1, a second hyperrecombination mutant plant was isolated called *sm22.* The original hyper-recombination phenotype of sm22 plant shows an enhancement of about 20- to 50- fold for homologous recombination in the reporter line. No other obvious phenotype was seen and the seed yield was normal. Sulfonamide selection in the second generation (T2) revealed a 2/1 or 3/1 segregation of resistant seedlings, thus showing that there is only 1 locus (or 2 closely related loci) with an active T-DNA inserted. However, the T2 recombination phenotype was even lower (less or same number of recombination events per plant) than in the wild type.

After HindIII digestion of T1 callus genomic DNA prepared essentially according to the method of the Nucleon Phytopure protocol and Plant DNA extraction kit (Amersham), plasmid rescue was applied as described in example 1, which gave rise to two independent

junction fragments. The first one corresponds to a single T-DNA insertion without deletion (left border, LB, junction sequenced) in the N-terminal region of a putative ATPase/helicase gene At3g57300, in antisense orientation. The second T-DNA inserted in a gene with no obvious relationship to homologous recombination (gb AF082176_1) and does not confer sulfonamide resistance. Six (T3) resistant families were analysed by PCR and Southern. Only one family contained some plants with the second insertion whereas all families have the helicase insertion site.

In subsequent generations, homozygous plants for the helicase insertion site were obtained. The homologous recombination frequency of heterozygous and homozygous plants for this insertion site was 80% and 20%, respectively, of the wild type level.

The predicted helicase gene (8kb genomic DNA) has about 20 exons encoding a protein of about 1489 amino acids. It is predicted to be an ATPase of the Swi2/Snf2 family, and contains several nuclear localization signals (NLS). The complete cDNA (4.8kb) was cloned in two steps. First, a public EST containing the 3' part was sequenced. Then the 5' part of the cDNA was amplified by RT-PCR on Col-0 (*Arabidopsis* Columbia ecotype, wild type) callus RNA (prepared with the Qiagen RNAeasy Plant Kit), using primers in the 5' untranslated region including a stop codon in frame with the predicted ATG (sm5UT) to make sure that the complete 5' part of the cDNA was amplified. The primer sequences were sm5UT: ctagaagcttttaaggat*TAA*gactctcc and for 3' primer: ctcgtatgtatccccttctcc.

The ATPase/helicase encoded by the gene (AGI: At3g57300) is the putative Arabidopsis ortholog of the yeast Ino80p/YGL150c protein (Ebbert et al. (1999), Shen et al. (2000). Homologs exist in yeast, budding yeast, *Drosophila* and human. These four homologues have several highly conserved regions including the six motifs of the SWI2/SNF2 helicase domain. Several NLS suggest a nuclear localization of the gene product.

The yeast homolog (Ebbert *et al.*,1999), INO80(=YGL150c), which is part of a big complex >1MDa (monomeric form is 171KDa), containing two essential helicases Rvb1p and Rvb2p, implicates these genes in homologous recombination in Eukaryotes (Cho et al. 2001; Jonsson et al. 2001; Wood et al. 2000). Human Rvb1p and Rvb2p are also known (Kanemaki 1999, Ikura et al. 2000, Shen et al. 2000). In *Arabidopsis thaliana* we found three genes closely related to Rvbs from other organisms (. The first one is the ortholog of yRvb1

and we named it AtRvb1. We found two counterparts for yRvb2 that we named AtRvb21 and AtRvb22. The three genes are expressed (RT-PCR) and some of them are positively regulated by genotoxic stress (UVc, bleomycin). For treatment with Bleomycin (BLM) 2 week-old Arabidopsis seedlings were placed under sterile conditions in liquid GM medium containing 10-6M of BLM (Sigma) or 100 ppm of MMS (Fluka, Switzerland). For UV-C irradiation (6000 ergs) 2 week-old seedlings were irradiated with light provided by a HNS 55W OFR lamp (Osram). After treatment, plants were harvested at several time points (30min, 1h, 4h and 12h) and RNA extracted as described above. Then semi-quantitative RT-PCR analysis was performed with the following primers AtIno80 (TGATGGATCTATCACCATCAG ggtgggattccaatcactttc) AtRvb1 (tttgatgggccaaatgatg cttccaaCCTAGGtgagatgtttcaacaaaatgtgc) AtRvb21 (tcaacagcaggacacaagg cccaatgCCTAGGaaatccgagttcaacatcctaatc) AtRvb22 (acaaaccagatatcagcacatgg aacaagtactcgctctcatgctc). In the sm22 background the steady state level of AtRvb21 and AtRvb22 was shown to be down-regulated using RT-PCR on RNA extracted as above mentioned.

This indicates that the components of the putative Arabidopsis Ino80 complex show co-regulation at the transcriptional level, supporting the use of Arabidopsis Rvb1, Rvb21 and Rvb22 to manipulate homologous recombination frequency in plants.

Example 6: AtRvb1 as positive regulator of homologous recombination.

As describe above (Example 5), the original recombination-up phenotype found in sm22 can be associated with an effect mediated by the Arabidopsis Rvb1 and 2 orthologs. Thus, AtRvb1 can be used as a positive regulator of homologous recombination.

Example 7: AtRvb21 as positive regulator of homologous recombination.

As describe above (Example 5), the original recombination-up phenotype found in sm22 can be associated with an effect mediated by the Arabidopsis Rvb1 and 2 orthologs. Thus, AtRvb21 can be used as a positive regulator of homologous recombination.

Example 8: AtRvb22 as positive regulator of homologous recombination.

As describe above (Example 5), the original recombination-up phenotype found in sm22 can be associated with an effect mediated by the Arabidopsis Rvb1 and 2 orthologs. Thus, AtRvb22 can be used as a positive regulator of homologous recombination.

Example 9: **At3g57290 as positive regulator of homologous recombination.**

In the *sm22* mutant (Example 5), the At3g57290p gene is potentially overexpressed by the 35S Enhancer/promoter. Over expression of this gene in the sm22 context or directly with a 35S promoter can be carried out to reproduce the original recombination-up phenotype. The phenotype was lost in the second generation (Example 5), at which point At3g57290 is not overexpressed any longer allowing a temporal ability to modulate homologous recombination.

All publications referred to herein are incorporated by reference as if each is referred to individually.

What is claimed is:

1. An isolated nucleic acid comprising a sequence having 98.5% or more identity with the sequence depicted in Figure 1C or Figure 1E.

2. The nucleic acid of claim 1, wherein said nucleic acid is DNA.

3. A vector comprising the nucleic acid if claim 2.

4. A host cell comprising the vector or nucleic acid of claim 3.

5. A polypeptide encoded by the isolated nucleic acid of claim 1.

6. An isolated nucleic acid comprising a sequence having 98.5% or more identity with the sequence depicted in Figure 5A.

7. The nucleic acid of claim 6, wherein said nucleic acid is DNA.

8. A vector comprising the nucleic acid of claim 7.

9. A host cell comprising the vector or nucleic acid of claim 8.

10. A polypeptide encoded by the isolated nucleic acid of claim 6.

11. A method for inducing homologous recombination in a cell, said method comprising modulating the expression or properties of one or more gene products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21 (At5g67630), Rvb22 (At3g49830) and At3g57290.

12. The method of claim 11, said method comprising increasing expression of said gene product.

13. The method of claim 12, said method comprising introducing a nucleic acid encoding said gene product into said cell operably linked to a promoter; and allowing transcription and translation of said gene in an amount sufficient to affect homologous recombination in said cell.

14. The method of claim 13, wherein said homologous recombination is somatic homologous recombination.

15. The method of claim 13, wherein said homologous recombination is meiotic homologous recombination.

16. The method of claim 13, wherein said promoter is an inducible promoter.

17. The method of claim 13, wherein said promoter is a tissue-specific promoter.

18. The method of claim 13, wherein said promoter is a constitutive promoter.

19. The method of claim 13, wherein said promoter is a meiosis-specific promoter.

20. A method of increasing gene targetting to a desired locus in a host cell, said method comprising introducing a desired gene into a host cell, modulating the expression or properties of one or more gene products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21 (At5g67630), Rvb22 (At3g49830) and At3g57290, or functional fragments, derivatives and homologues thereof in said host cell, and detecting integration of said desired gene at a selected locus in the genome of said host cell.

Abstract

The present invention relates to nucleic acids encoding polypeptides involved in homologous recombination, as well as vectors and host cells comprising the nucleic acids and polypeptides encoded by the nucleic acids. Also provided are methods for inducing somatic and/or meiotic homologous recombination in a cell, comprising modulating the expression or properties of one or more gene products selected from the group consisting of MBM17.5, MBM17.6, osMBM17.5, zmMBM17.5, AtIno80, At3g57300, Rvb1 (At5g22330), Rvb21 (At5g67630), Rvb22 (At3g49830) and At3g57290, their homologues, fragments or derivatives. In particular, the methods can be used to increase gene targetting.

Figure 1A

>predicted cDNA of mbm17.5

atggcggaaaatacggccagccatagaagaaaacctcggagcttgaacgatcgtcactacagtatcctccaggatctttctgcg
cctcctagacagcctccctcttcttctcatggagaagatgaagagacgaagaagtccatgattaagcttgctggacgacgtcgtct
ttgcaaggccttgccaaaggaagacgaagctgatggatatgacgatcctgatttggttgatttctattccccagttaaaggagaga
catcactagacagcgctggaattgggaacaaattcacatcttgggatgaatcaaaggaagctaacactgagctggctggcgag
cctaacttttcgataatcacagacttttgttcgccctcacctcagttgaagcaaaaagaggaaatgcaaggtgatggaggaagga
acgagatcatgggtattttggatgatttgacctctaagcttgggacaatgtcgattcagaagaagaaggatagccaaagcaatga
ttttgatgcatgtggagtgaagagccaggttgataaatttgattttgaggatgccaaatcctcattttccttgctatcggatctatctaag
tcctcaccagatgtggttaccacatataatgctggcgttaatagtatcaaggacaagcaaggcaaatctggttttgccatccggga
agagcaaactagtaaggaattttcaagggaatgggaagaaagaatttcgaatgttggaaagcaaaactcatattctggtcggc
actttgacgataactctgaagataataggcagggatacaatcttgaccgtgggaagagccaatgcaaggaagtcgaccaaagt
atgaagacgaccaggcacatagaggtaagtgagaagataagaacagtcggaaggtctaatgctgccaagctaagagactta
gacgaggatgatgatgatgatgactgtctcattttgtccgggaaaaaggcggctgaaatgaaaattaataagccagctcggtctt
ataacgccaaaagacatggttatgatgagagatcgttggaagatgaagggtctatcactttaactggcctcaatttgtcttacacat
tgcctggaaagattgcaacaatgttatatccacatcagagggaagggttgaattggctttggtcattgcatacccaagggaaagg
tggaatacttggagatgatatgggtttaggtaaaactatgcagatttgtagttttcttgctggtttattccactccaaattgatcaagcgt
gctctggtagtggccccaaaaaccttgctgcctcactggatgaaagaattagctaccgtgggactttcacaaatgactagggaat
actacggtacttctacgaaagcccgggaatatgatctccaccacattctgcagggtaaaggtattcttctaacaacctatgatattgt
gcggaacaatacaaaggctttgcaaggtgacgaccattatactgatgaggatgatgaagatggaaacaaatgggactacatg
attctggacgagggacatcttattaagaaccccaacacacaaagggcgaagagtttgcttgagatcccaagttctcaccgtattat
aataagtggtacaccaatccagaacaatctcaaggtattattgtctatgacatttaacgttgctgccctgggttactcggtgacaag
aattggtaaacatatcctccctataattttgtcagagtacatttaccagctctgtattactaaggctttaagctttaacaggtttaagcag
aattatgagcattacattcttcgtggaactgacaaaaatgctactgatagagaacagaggataggctcaacagtagcaaagaa
cttgagggagcatattcaacctttcttcttgcggcgccttaagagtgaagtcttcggtgatgatggtgcaacctccaaactttcgaag
aaggacgaaattgttgtatggttacggttaacagcttgccagaggcaattatatgaagctttcttaaacagtgaaattgttctgtcag
cttttgatggttcacctctagcagctctaacgattctgaagaaaatatgtgaccacccgcttctcttaactaagagggctgctgagga
tgtccttgaaggaatggattcaacattaacacaagaagaagcaggcgtggctgagagattggctatgcatatagcggacaatgt
ggatacagatgattttcagaccaagaatgacagtatctcttgcaaattgtcatttatcatgtcgctactggaatttcaagaaggtcat
gtggctcctatatttctcttgacttctcaagttggtggtctcggccttactctgactaaggcagaccgtgtgattgtggtggaccctgcct
ggaatccaagcacggacaaccagagtgttgatcgagcatatagaattgggcagacaaaggatgtcatcgtatataggttaatg
acctcagcaactgttgaagaaaagatatacagaaagcaggtatacaagggaggattgtttaaaactgcaactgagcataaag
aacaaatccgctacttcagccagcaggaccttcgagaacttttagtcttcccaagggaggctttgatgtttcacctacacaacagc

aactatacgaagagcactataaccaaatcaaactagatgaaaaactggaatcccattgtaaagtttctcgaaacccttggtatag
ctggagttagccaccatagcttacttttctccaagacagctcctattcaagcgatacagaaagatgaagaagaacaaataaggg
ctgactatgctttcaagccaaaggatgtgaatttggacaagagaatcaacatttccccagtcgatgacaaggaattgtcagaaag
cgtaattaaagcaagactcaatcgtttgacgatgctattacaaaacaagggtacggtctcaaggctacctgatggaggggcaaa
aatccagaagcagattgctgaattgactcgagaactgaaagacatgaaagcagcagaaaggatcaacatgcctcaagttattg
acttggaggaggatataagtcggaagatgcaaaaaggattgaatctgtag

Figure 1B

>predicted protein sequence of MBM17.5

MAENTASHRRKPRSLNDRHYSILQDLSAPPRQPPSSSHGEDEETKKSMIKLAGRRRLCKAL
PKEDEADGYDDPDLVDFYSPVKGETSLDSAGIGNKFTSWDESKEANTELAGEPNFSIITDFC
SPSPQLKQKEEMQGDGGRNEIMGILDDLTSKLGTMSIQKKKDSQSNDFDACGVKSQVDKF
DFEDAKSSFSLLSDLSKSSPDVVTTYNAGVNSIKDKQGKSGFAIREEQTSKEFSREWEERIS
NVGKQNSYSGRHFDDNSEDNRQGYNLDRGKSQCKEVDQSMKTTRHIEVSEKIRTVGRSNA
AKLRDLDEDDDDDDCLILSGKKAAEMKINKPARSYNAKRHGYDERSLEDEGSITLTGLNLSY
TLPGKIATMLYPHQREGLNWLWSLHTQGKGGILGDDMGLGKTMQICSFLAGLFHSKLIKRA
LVVAPKTLLPHWMKELATVGLSQMTREYYGTSTKAREYDLHHILQGKGILLTTYDIVRNNTK
ALQGDDHYTDEDDEDGNKWDYMILDEGHLIKNPNTQRAKSLLEIPSSHRIIISGTPIQNNLKV
LLSMTFNVAALGYSVTRIGKHILPIILSEYIYQLCITKALSFNRFKQNYEHYILRGTDKNATDRE
QRIGSTVAKNLREHIQPFFLRRLKSEVFGDDGATSKLSKKDEIVVWLRLTACQRQLYEAFLN
SEIVLSAFDGSPLAALTILKKICDHPLLLTKRAAEDVLEGMDSTLTQEEAGVAERLAMHIADNV
DTDDFQTKNDSISCKLSFIMSLLEFQEGHVAPIFLLTSQVGGLGLTLTKADRVIVVDPAWNPS
TDNQSVDRAYRIGQTKDVIVYRLMTSATVEEKIYRKQVYKGGLFKTATEHKEQIRYFSQQDL
RELFSLPKGGFDVSPTQQQLYEEHYNQIKLDEKLESHVKFLETLGIAGVSHHSLLFSKTAPIQ
AIQKDEEEQIRADYAFKPKDVNLDKRINISPVDDKELSESVIKARLNRLTMLLQNKGTVSRLP
DGGAKIQKQIAELTRELKDMKAAERINMPQVIDLEEDISRKMQKGLNL

Figure 1C

>full length cDNA of mbm17.5

```
   1 ATGGCGGAAA ATACGGCCAG CCATAGAAGA AAACCTCGGA GCTTGAACGA
  51 TCGTCACTAC AGTATCCTCC AGGATCTTTC TGCGCCTCCT AGACAGCCTC
 101 CCTCTTCTTC TCATGGAGAA GATGAAGAGA CGAAGAAGTC CATGATTAAG
 151 CTTGCTGGAC GACGTCGTCT TTGCAAGGCC TTGCCAAAGG AAGACGAAGC
 201 TGATGGATAT GACGATCCTG ATTTGGTTGA TTTCTATTCC CCAGTTAAAG
 251 GAGAGACATC ACTAGACAGC GCTGGAATTG GGAACAAATT CACATCTTGG
 301 GATGAATCAA AGGAAGCTAA CACTGAGCTG GCTGGCGAGC CTAACTTTTC
 351 GATAATCACA GACTTTTGTT CGCCCTCACC TCAGTTGAAG CAAAAAGAGG
 401 AAATGCAAGG TGATGGAGGA AGGAACGAGA TCATGGGTAT TTTGGATGAT
```

451 TTGACCTCTA AGCTTGGGAC AATGTCGATT CAGAAGAAGA AGGATAGCCA
501 AAGCAATGAT TTTGATGCAT GTGGAGTGAA GAGCCAGGTT GATAAATTTG
551 ATTTTGAGGA TGCCAAATCC TCATTTTCCT TGCTATCGGA TCTATCTAAG
601 TCCTCACCAG ATGTGGTTAC CACATATAAT GCTGGCGTTA ATAGTATCAA
651 GGACAAGCAA GGCAAATCTG GTTTTGCCAT CCGGGAAGAG CAAACTAGTA
701 AGGAATTTTC AAGGGAATGG GAAGAAAGAA TTTCGAATGT TGGAAAGCAA
751 AACTCATATT CTGGTCGGCA CTTTGACGAT AACTCTGAAG ATAATAGGCA
801 GGGATACAAT CTTGACCGTG GGAAGAGCCA ATGCAAGGAA GTCGACCAAA
851 GTATGAAGAC GACCAGGCAC ATAGAGGTAA GTGAGAAGAT AAGAACAGTC
901 GGAAGGTCTA ATGCTGCCAA GCTAAGAGAC TTAGACGAGG ATGATGATGA
951 TGATGACTGT CTCATTTTGT CCGGGAAAAA GGCGGCTGAA ATGAAAATTA
1001 ATAAGCCAGC TCGGTCTTAT AACGCCAAAA GACATGGTTA TGATGAGAGA
1051 TCGTTGGAAG ATGAAGGGTC TATCACTTTA ACTGGCCTCA ATTTGTCTTA
1101 CACATTGCCT GGAAAGATTG CAACAATGTT ATATCCACAT CAGAGGGAAG
1151 GGTTGAATTG GCTTTGGTCA TTGCATACCC AAGGGAAAGG TGGAATACTT
1201 GGAGATGATA TGGGTTTAGG TAAAACTATG CAGATTTGTA GTTTTCTTGC
1251 TGGTTTATTC CACTCCAAAT TGATCAAGCG TGCTCTGGTA GTGGCCCCAA
1301 AAACCTTGCT GCCTCACTGG ATGAAAGAAT TAGCTACCGT GGGACTTTCA
1351 CAAATGACTA GGGAATACTA CGGTACTTCT ACGAAAGCCC GGGAATATGA
1401 TCTCCACCAC ATTCTGCAGG GTAAAGGTAT TCTTCTAACA ACCTATGATA
1451 TTGTGCGGAA CAATACAAAG GCTTTGCAAG GTGACGACCA TTATACTGAT
1501 GAGGATGATG AAGATGGAAA CAAATGGGAC TACATGATTC TGGACGAGGG
1551 ACATCTTATT AAGAACCCCA ACACACAAAG GGCGAAGAGT TTGCTTGAGA
1601 TCCCAAGTTC TCACCGTATT ATAATAAGTG GTACACCAAT CCAGAACAAT
1651 CTCAAGGAAC TGTGGGCTCT CTTCAACTTC AGCTGCCCTG GGTTACTCGG
1701 TGACAAGAAT TGGTTTAAGC AGAATTATGA GCATTACATT CTTCGTGGAA
1751 CTGACAAAAA TGCTACTGAT AGAGAACAGA GGATAGGCTC AACAGTAGCA
1801 AAGAACTTGA GGGAGCATAT TCAACCTTTC TTCTTGCGGC GCCTTAAGAG
1851 TGAAGTCTTC GGTGATGATG GTGCAACCTC CAAACTTTCG AAGAAGGACG
1901 AAATTGTTGT ATGGTTACGG TTAACAGCTT GCCAGAGGCA ATTATATGAA
1951 GCTTTCTTAA ACAGTGAAAT TGTTCTGTCA GCTTTTGATG GTTCACCTCT
2001 AGCAGCTCTA ACGATTCTGA AGAAAATATG TGACCACCCG CTTCTCTTAA
2051 CTAAGAGGGC TGCTGAGGAT GTCCTTGAAG GAATGGATTC AACATTAACA
2101 CAAGAAGAAG CAGGCGTGGC TGAGAGATTG GCTATGCATA TAGCGGACAA
2151 TGTGGATACA GATGATTTTC AGACCAAGAA TGACAGTATC TCTTGCAAAT
2201 TGTCATTTAT CATGTCGCTA CTGGAAAATT TAATTCCAGA GGGGCACCGT
2251 GTTCTAATCT TCTCCCAGAC ACGCAAGATG CTTAATCTCA TTCAGGATTC
2301 TCTTACCTCC AACGGTTATA GTTTCTTGCG AATTGATGGT ACAACAAAAG
2351 CCCCTGACAG ATTGAAGACT GTTGAAGAAT TTCAAGAAGG TCATGTGGCT
2401 CCTATATTTC TCTTGACTTC TCAAGTTGGT GGTCTCGGCC TTACTCTGAC
2451 TAAGGCAGAC CGTGTGATTG TGGTGGACCC TGCCTGGAAT CCAAGCACGG
2501 ACAACCAGAG TGTTGATCGA GCATATAGAA TTGGGCAGAC AAAGGATGTC
2551 ATCGTATATA GGTTAATGAC CTCAGCAACT GTTGAAGAAA AGATATACAG
2601 AAAGCAGGTA TACAAGGGAG GATTGTTTAA AACTGCAACT GAGCATAAAG
2651 AACAAATCCG CTACTTCAGC CAGCAGGACC TTCGAGAACT TTTTAGTCTT
2701 CCCAAGGGAG GCTTTGATGT TTCACCTACA CAACAGCAAC TATACGAAGA
2751 GCACTATAAC CAAATCAAAC TAGATGAAAA ACTGGAATCC CATGTAAAGT
2801 TTCTCGAAAC CCTTGGTATA GCTGGAGTTA GCCACCATAG CTTACTTTTC
2851 TCCAAGACAG CTCCTATTCA AGCGATACAG AAAGATGAAG AAGAACAAAT
2901 AAGGAGAGAA ACAGCATTGC TCTTGGGACG CGCATCAGCA AGTATTTCAC
2951 AAGACACCGT CATCAATGGG GCTGACTATG CTTTCAAGCC AAAGGATGTG

3001 AATTTGGACA AGAGAATCAA CATTTCCCCA GTCGATGACA AGGAATTGTC
3051 AGAAAGCGTA ATTAAAGCAA GACTCAATCG TTTGACGATG CTATTACAAA
3101 ACAAGGGTAC GGTCTCAAGG CTACCTGATG GAGGGGCAAA AATCCAGAAG
3151 CAGATTGCTG AATTGACTCG AGAACTGAAA GACATGAAAG CAGCAGAAAG
3201 GATCAACATG CCTCAAGTTA TTGACTTGGA GGAGGATATA AGTCGGAAGA
3251 TGCAAAAAGG ATTGAATCTG TAG

Figure 1D

>protein sequence of MBM17.5

MAENTASHRRKPRSLNDRHYSILQDLSAPPRQPPSSSHGEDEETKKSMIK 50
LAGRRRLCKALPKEDEADGYDDPDLVDFYSPVKGETSLDSAGIGNKFTSW 100
DESKEANTELAGEPNFSIITDFCSPSPQLKQKEEMQGDGGRNEIMGILDD 150
LTSKLGTMSIQKKKDSQSNDFDACGVKSQVDKFDFEDAKSSFSLLSDLSK 200
SSPDVVTTYNAGVNSIKDKQGKSGFAIREEQTSKEFSREWEERISNVGKQ 250
NSYSGRHFDDNSEDNRQGYNLDRGKSQCKEVDQSMKTTRHIEVSEKIRTV 300
GRSNAAKLRDLDEDDDDDDCLILSGKKAAEMKINKPARSYNAKRHGYDER 350
SLEDEGSITLTGLNLSYTLPGKIATMLYPHQREGLNWLWSLHTQGKGGIL 400
GDDMGLGKTMQICSFLAGLFHSKLIKRALVVAPKTLLPHWMKELATVGLS 450
QMTREYYGTSTKAREYDLHHILQGKGILLTTYDIVRNNTKALQGDDHYTD 500
EDDEDGNKWDYMILDEGHLIKNPNTQRAKSLLEIPSSHRIIISGTPIQNN 550
LKELWALFNFSCPGLLGDKNWFKQNYEHYILRGTDKNATDREQRIGSTVA 600
KNLREHIQPFFLRRLKSEVFGDDGATSKLSKKDEIVVWLRLTACQRQLYE 650
AFLNSEIVLSAFDGSPLAALTILKKICDHPLLLTKRAAEDVLEGMDSTLT 700
QEEAGVAERLAMHIADNVDTDDFQTKNDSISCKLSFIMSLLENLIPEGHR 750
VLIFSQTRKMLNLIQDSLTSNGYSFLRIDGTTKAPDRLKTVEEFQEGHVA 800
PIFLLTSQVGGLGLTLTKADRVIVVDPAWNPSTDNQSVDRAYRIGQTKDV 850
IVYRLMTSATVEEKIYRKQVYKGGLFKTATEHKEQIRYFSQQDLRELFSL 900
PKGGFDVSPTQQQLYEEHYNQIKLDEKLESHVKFLETLGIAGVSHHSLLF 950
SKTAPIQAIQKDEEEQIRRETALLLGRASASISQDTVINGADYAFKPKDV 1000
NLDKRINISPVDDKELSESVIKARLNRLTMLLQNKGTVSRLPDGGAKIQK 1050
QIAELTRELKDMKAAERINMPQVIDLEEDISRKMQKGLNL*

Figure 1E

>over-expressed transcript of mbm17.5 in mutant hw17

1 AGAGGACAGG GTACCCGGGG ATCAGATTGT CGTTTCCCGC CTTCAGTTTA
51 AACTATCAGT GTTTGAATTG AAGTATTCTT TATATGTTAC GCATGGAATT
101 TTCAGGATTC TCTTACCTCC AACGGTTATA GTTTCTTGCG AATTGATGGT
151 ACAACAAAAG CCCCTGACAG ATTGAAGACT GTTGAAGAAT TTCAAGAAGG
201 TCATGTGGCT CCTATATTTC TCTTGACTTC TCAAGTTGGT GGTCTCGGCC
251 TTACTCTGAC TAAGGCAGAC CGTGTGATTG TGGTGGACCC TGCCTGGAAT
301 CCAAGCACGG ACAACCAGAG TGTTGATCGA GCATATAGAA TTGGGCAGAC
351 AAAGGATGAC ATCGTATATA GGTTAATGTC CTCAGCAACT GTTGAAGAAA
401 AGATATACAG AAAGCAGGTA TACAAGGGAG GATTGTTTAA AACTGCAACT
451 GAGCATAAAG AACAAACCCG CTACTTCAGC CAGCAGGACC TTCGAGAACT
501 TTTTAGTCTT CCCAAGGGAG GCTTTGATGT TTCACCTACA CAACAGCAAC
551 TATACGAAGA GCACTATAAC CGAATCAAAC TAGATGAAAA ACTGGAATCC
601 CATGTAAAGT TTCTCGAAAC CCTTGGTATA GCTGGAGTTA GCCACCATAG
651 CTTACTTTTC TCCAAGACAG CTCCTATTCA AGCGATACAG AAAGATGAAG
701 AAGAACAAAT AAGGAGAGAA ACAGCATTGC TCTTGGGACG CGCATCAGCA
751 AGTATTTCAC AAGACACCGT CATCAATGGG GCTGACTATG CTTTCAAGCC

801 AAAGGATGTG AATTTGGACA AGAGAATCAA CATTTCCCCA GTCGATGACA
851 AGGAATTGTC AGAAAGCGTA ATTAAAGCAA GACTCAATCG TTTGACGATG
901 CTATTACAAA ACAAGGGTAC GGTCTCAAGG CTACCTGATG GAGGGGCAAA
951 AATCCAGAAG CAGATTGCTG AATTGACTCG AGAACTGAAA GACATGAAAG
1001 CAGCAGAAAG GATCAACATG CCTCAAGTTA TTGACTTGGA GGAGGATATA
1051 AGTCGGAAGA TGCAAAAAGG ATTGAATCTG TAGAGTAAGA TACAAGTCAA
1101 GATGCAAGAA ATGCAAACGA CCATCATTGC AACACTTGTG GTTTTTTTTT
1151 GTTCCTTATC TAATTTGGTT TGGTTGAATT GGTAAGTCAA TTACCATATG
1201 ACTTGCTGCA AAAAAAAAAA AAAAAAA

Fig.2: sequences related to mbm17.6

Figure 2A.

>predicted cDNA of mbm17.6 (DNA polymerase III)

1 ATGAATAGAT CCGGTATTTC CAAAAAGCGA CCGCCTCCTT CGAATACCCC
51 ACCACCGGCG GGTAAGCATC GAGCCACTGG TGATTCAACA CCATCTCCGG
101 CCATCGGAAC CCTAGATGAT GAATTTATGA TGGAAGAGGA CGTGTTTCTG
151 GACGAAACTC TCTTGTACGG CGACGAAGAT GAGGAATCCC TAATCCTCCG
201 TGACATTGAG GAGCGTGAAT CGCGTTCCTC GGCTTGGGCT CGACCTCCGC
251 TCTCCCCCGC GTATCTCTCG AATTCACAGA TTTTCCAACA ATTGGAGATT
301 GACTCTATAA TCGCGGAGAG TCATAAGGAG CTGTTACCGG GTTCCTCAGG
351 GCAAGCTCCA ATCATTAGGA TGTTTGGGGT TACCAGAGAA GGTAACAGTG
401 TGTGTTGCTT TGTTCATGGA TTTGAGCCAT ACTTTTACAT TGCTTGCCCT
451 CCTGGAATGG GGCCAGACGA TATTTCTAAT TTCCATCAGA GTCTTGAGGG
501 AAGGATGAGG GAATCCAATA AAAATGCCAA GGTCCCGAAA TTTGTTAAAC
551 GTATAGAAAT GGTGCAGAAA AGAAGCATTA TGTATTACCA ACAGCAAAAA
601 TCCCAAACTT TTCTGAAGAT TACAGTTGCA TTGCCGACTA TGGTGGCAAG
651 CTGTCGCGGC ATCCTTGATA GAGGCCTACA AATTGATGGA TTGGGTATGA
701 AGAGCTTCCA GACATATGAA AGCAATATTC TTTTCGTTCT CCGTTTCATG
751 GTTGATTGTG ATATTGTCGG AGGAAATTGG ATTGAAGTAC CTACTGGGAA
801 GTATAAGAAA AATGCAAGAA CTTTGTCATA CTGCCAATTG GAGTTCCATT
851 GCCTGTACTC AGATCTAATC AGTCATGCTG CAGAAGGTGA ATACTCAAAA
901 ATGGCTCCAT TCCGTGTACT AAGTTTCGAT ATTGAGTGTG CAGGTCGTAA
951 AGGACATTTT CCGGAAGCTA AGCATGATCC TGTAATCCAG ATAGCGAACC
1001 TTGTTACTCT TCAGGGAGAG GATCACCCAT TTGTACGCAA TGTCATGACT
1051 CTTAAGTCAT GTGCTCCAAT CGTAGGCGTA GATGTCATGT CTTTTGAAAC
1101 AGAAAGAGAG GTCTTACTAG CTTGGAGGGA TTTGATTCGT GATGTTGATC
1151 CTGATATCAT CATTGGTTAT AACATCTGCA AATTCGATTT ACCTTATCTG
1201 ATTGAGAGAG CTGCTACACT GGGAATAGAG GAATTTCCTC TTCTTGGTCG
1251 TGTAAAGAAC AGTAGGGTCC GGGTCAGGGA CTCAACATTT TCATCAAGAC
1301 AACAAGGAAT AAGAGAAAGT AAAGAGACCA CAATTGAAGG AAGATTTCAG
1351 TTTGACCTTA TTCAGGCAAT ACACAGAGAC CACAAATTAA GTTCTTATTC
1401 GCTGAATTCT GTCTCAGCTC ACTTTCTTTC CGAGCAGAAA GAAGATGTCC
1451 ACCATTCTAT AATAACTGAT CTCCAGAATG GGAATGCGGA AACCAGGAGG
1501 CGTCTTGCTG TTTATTGTTT GAAGGATGCA TATCTTCCTC AGAGGCTTCT
1551 GGACAAACTG ATGTTTATAT ATAATTATGT CGAAATGGCT CGTGTAACTG
1601 GTGTCCCTAT TTCATTTCTT CTTGCGAGAG GACAGAGTAT CAAGGTTTTA
1651 TCTCAGCTTC TTAGGAAAGG CAAACAGAAA AATCTGGTTC TTCCAAATGC
1701 TAAACAGTCA GGGTCCGAAC AAGGAACTTA TGAAGGCGCA ACTGTTTTAG
1751 AAGCAAGAAC AGGTTTCTAT GAAAAGCCAA TTGCAACTTT GGATTTTGCT

1801 TCACTGTACC CGTCAATTAT GATGGCATAT AATCTGTGCT ACTGCACCTT
1851 GGTGACACCT GAAGATGTAC GCAAACTGAA TCTTCCACCT GAACATGTCA
1901 CTAAAACTCC ATCAGGGGAA ACATTTGTTA AGCAAACTTT GCAAAAGGGT
1951 ATACTTCCAG AAATTCTCGA AGAGCTTCTT ACTGCCCGTA AGAGAGCTAA
2001 AGCAGATTTA AAGGAGGCTA AGGATCCCCT TGAGAAGGCT GTTTTAGATG
2051 GTAGACAGTT AGCGTTGAAG ATCAGTGCAA ATTCTGTCTA CGGGTTTACG
2101 GGAGCCACTG TTGGGCAGTT ACCATGCTTA GAAATATCCT CGAGTGTAAC
2151 TAGCTATGGT CGTCAGATGA TTGAACAAAC AAAGAAACTT GTTGAAGACA
2201 AATTCACAAC ACTGGGAGGG TATCAATACA ATGCAGAGGT CATTTATGGA
2251 GACACGGATT CAGTCATGGT GCAATTTGGA GTATCGGATG TAGAAGCTGC
2301 GATGACCTTG GGGAGGGAAG CTGCAGAACA CATTAGTGGA ACTTTTATCA
2351 AACCCATCAA ATTGGAGTTT GAAAAGGTCT ATTTCCCATA TCTTCTCATT
2401 AACAAGAAGA GGTATGCTGG TTTGCTATGG ACAAATCCTC AACAGTTTGA
2451 CAAAATGGAC ACCAAAGGAA TCGAGACAGT ACGAAGGGAT AATTGTTTAC
2501 TGGTTAAGAA CCTCGTGACT GAGAGTCTTA ACAAAATACT TATTGATAGA
2551 GATGTTCCAG GGGCAGCTGA AAATGTCAAG AAAACCATTT CGGATCTTCT
2601 CATGAACCGT ATTGACTTGT CACTTTTGGT GATTACTAAG GGTCTAACGA
2651 AAACAGGAGA TGATTATGAA GTTAAATCAG CTCATGGTGA ACTTGCTGAA
2701 CGCATGCGTA AGAGGGATGC TGCTACAGCG CCAAATGTTG GAGATCGAGT
2751 ACCGTATGTT ATCATAAAAG CTGCTAAAGG TGCCAAGGCT TATGAACGAT
2801 CAGAAGATCC AATCTACGTG CTACAGAATA ATATCCCTAT AGACCCAAAT
2851 TACTACTTGG AGAATCAGAT TAGCAAGCCA CTTCTTAGGA TTTTTGAGCC
2901 AGTCCTGAAA AATGCTAGCA AGGAGCTTCT CCATGGAAGT CACACGAGGT
2951 CAATATCAAT CACTACTCCT TCAAACAGCG GTATAATGAA GTTTGCTAAA
3001 AAACAACTGA GCTGTGTTGG CTGCAAAGTT CCGATCAGGT ACTTTGTGCA
3051 ATGGAACACT ATGCGCAAGT TGCAAGGGAA GAGAAGCCGA GTTATATTGC
3101 AAAAACGTGT CTCAAGGTAT GCTGCCTGGC TGAGCTTGAA GAGGTTTTTG
3151 GGAGGCTGTG GACACAGTGC CAGGAGTGTC AAGGCTCTCT TCATCAAGAT
3201 GTCTTGTGCA CCAGTCGAGA TTGTCCAATA TTTTACCGGA GAATGA

Figure 2B

>predicted protein sequence of MBM17.6 (DNA polymerase III)

MNRSGISKKRPPPSNTPPPAGKHRATGDSTPSPAIGTLDDEFMMEEDVFL 50
DETLLYGDEDEESLILRDIEERESRSSAWARPPLSPAYLSNSQIFQQLEI 100
DSIIAESHKELLPGSSGQAPIIRMFGVTREGNSVCCFVHGFEPYFYIACP 150
PGMGPDDISNFHQSLEGRMRESNKNAKVPKFVKRIEMVQKRSIMYYQQQK 200
SQTFLKITVALPTMVASCRGILDRGLQIDGLGMKSFQTYESNILFVLRFM 250
VDCDIVGGNWIEVPTGKYKKNARTLSYCQLEFHCLYSDLISHAAEGEYSK 300
MAPFRVLSFDIECAGRKGHFPEAKHDPVIQIANLVTLQGEDHPFVRNVMT 350
LKSCAPIVGVDVMSFETEREVLLAWRDLIRDVDPDIIIGYNICKFDLPYL 400
IERAATLGIEEFPLLGRVKNSRVRVRDSTFSSRQQGIRESKETTIEGRFQ 450
FDLIQAIHRDHKLSSYSLNSVSAHFLSEQKEDVHHSIITDLQNGNAETRR 500
RLAVYCLKDAYLPQRLLDKLMFIYNYVEMARVTGVPISFLLARGQSIKVL 550
SQLLRKGKQKNLVLPNAKQSGSEQGTYEGATVLEARTGFYEKPIATLDFA 600

SLYPSIMMAYNLCYCTLVTPEDVRKLNLPPEHVTKTPSGETFVKQTLQKG 650
ILPEILEELLTARKRAKADLKEAKDPLEKAVLDGRQLALKISANSVYGFT 700
GATVGQLPCLEISSSVTSYGRQMIEQTKKLVEDKFTTLGGYQYNAEVIYG 750
DTDSVMVQFGVSDVEAAMTLGREAAEHISGTFIKPIKLEFEKVYFPYLLI 800
NKKRYAGLLWTNPQQFDKMDTKGIETVRRDNCLLVKNLVTESLNKILIDR 850
DVPGAAENVKKTISDLLMNRIDLSLLVITKGLTKTGDDYEVKSAHGELAE 900
RMRKRDAATAPNVGDRVPYVIIKAAKGAKAYERSEDPIYVLQNNIPIDPN 950
YYLENQISKPLLRIFEPVLKNASKELLHGSHTRSISITTPSNSGIMKFAK 1000
KQLSCVGCKVPIRYFVQWNTMRKLQGKRSRVILQKRVSRYAAWLSLKRFL 1050
GGCGHSARSVKALFIKMSCAPVEIVQYFTGE*

Figure 3: Osmbm17.5

>EST clone RICS1367A, Oryza sativa homolog of mbm17.5, partial sequence

1 AGGAACTTTC AGTTGTGAGC CTCAAAGATA AGATCAGAGA CTACTCTGGT
51 CCCAATGCAA ATGCTCGCAA CTATGAGCTT AAATATGCCT TCAAGGAGGG
101 TGGAATCCTT TTAACAACAT ATGACATTGT TCGAAACAAT TTCAAGATGA
151 TAAAAGGCAA CTTCACCAAT GATTTTGATG ACGAGGAAGA AACATTATGG
201 AACTATGTTA TTCTTGATGA GGGGCATATT ATCAAGAATC CAAAGACTCA
251 GAGGGCTCAA AGTCTATTTG AAATACCCTG TGCACATCGT ATTGTCATCA
301 GTGGAACACC CATACAAAAT AACTTGAAGG AAATGTGGGC TCTGTTTTAT
351 TTCTGTTGCC CAGAAGTCTT GGGTGATAAG GAGCAGTTCA AAGCAAGGTA
401 TGAGCACGCT ATCATTCAAG GAAATGACAA GAATGCTACC AATCGACAAA
451 AGCACATAGG CTCAAATGTA GCAAAGGAAT TAAGAGAACG GATAAAGCCA
501 TACTTTTTGC GACGCATGAA GAATGAAGTG TTTCTTGATA GCGGCACGGG
551 AGAAGATAAA AAGCTTGCTA AGAAGAATGA GCTAATTATC TGGCTGAAAT
601 TAACATCTTG CCAGAGGCAA TTATATGAAG CTTTTCTTAA CAGTGAACTA
651 GTTCATTCAT CAATGCAAGG GTCACCCTTG GCCGCAATCA CGATATTGAA
701 GAAAATATGT GATCATCCGC TGTTGTTGAC TAAGAAAGCT GCTGAGGGTG
751 TTTTGGAAGG CATGGATGCG ATGTTAAATA ATCAAGAAAT GGGAATGGTT
801 GAGAAAATGG CCATGAACCT TGCAGATATG GCTCATGATG ATGATGACGT
851 TGAATTGCAA GTTGGTCAGG ATGTCTCGTG CAAGTTATCT TTTATGATGT
901 CCTTGCTCCA AAATCTTGTT AGCGAGGGAC ACAACGTCTT AATCTTCTCG
951 CAAACTCGTA AAATGCTAAA CATTATTCAG GAGGCTATAA TATTAGAAGG
1001 CTATAAGTTT TTGCGCATTG ATGGTACCAC CAAGATTTCT GAGAGGGAAA
1051 GGATTGTGAA GGACTTCCAA GAGGGTCCTG GAGCTCCAAT ATTTTTGCTG
1101 ACCACACAAG TTGGTGGGCT TGGACTTACA CTCACCAAGG CAGCTCGTGT
1151 CATAGTAGTT GATCCTGCTT GGAATCCAAG TACGGACAAT CAAAGTGTTG
1201 ATCGTGCTTA TCGAATTGGG CAGATGAAAG ATGTCATCGT ATACCGCCTT
1251 ATGACATCTG GAACCATCGA AGAAAAGATA TACAAATTGC AGGTCTTCAA
1301 GGGGGCTCTG TTTAGGACAG CTACAGAGCA CAAAGAACAA ACTCGTTATT
1351 TCAGCAAGAG GGATATTCAA GAGCTTTTCA GTCTGCCTGA GCAAGGTTTT
1401 GATGTTTCGC TGACACAAAA GCAATTGCAA GAAGAGCATG GACACCAACT
1451 TGTGATGGAC GACTCCTTGA GGAAGCATAT ACAATTCCTG GAGCAACAAG
1501 GCATAGCGGG CGTGAGCCAT CACAGCCTTC TGTTTTCTAA GACAGCAATC
1551 TTACCTACAC TGAATGATAA TGATGGTTTG GACAGTCGTC GAGCTATGCC
1601 AATGGCCAAG CACTACTACA AGGGAGCCTC ATCTGACTAT GTTGCCAATG

1651 GTGCTGCCTA TGCGATGAAG CCAAAAGAGT TCATTGCTCG AACATACTCC
1701 CCGAACAGCA CAAGCACAGA AAGTCCTGAG GAAATCAAGG CCAAAATCAA
1751 CCGGCTTTCG CAAACCCTTG CAAACACGGT GCTTGTGGCG AAGCTACCAG
1801 ATCGTGGAGA CAAGATAAGG AGGCAGATAA ATGAGCTGGA CGAAAAGCTG
1851 ACCGTGATCG AGTCTTCTCC GGAGCCATTG GAGAGGAAGG GTCCAACGGA
1901 AGTAATCTGC TTGGATGATC TGTCTGTCTA GTGTAGGGCA TGTCTGTTTC
1951 TTTTGCTTAA ATTCCATGCT TGCATGCTAG TAGTCACTAA GGCGTGACAT
2001 TTTGCATGCT ACTTGTACTA ATTGTGACGA CCACGGAACG GAACACATGC
2051 TGATCTCGGG TGCCTCTTAG GCTTGTGTCT GAGAGGAGAA AAAGAGAATA
2101 TTGACCAAAA AAAAAAAAAA

Fig 4: zmmbm17.5
>EST clone 603011H11, Zea mays homolog of mbm17.5, partial sequence
1 GAGTGGGACA ACCAGGACGA CGGTGAAAGC ATACTCGACA TCCTAGACGA
51 CCTCACCACA CGATTTGACT CTCTATCCGT CCAGAAGCCC AGCACCGCCG
101 CGAGGTCCAG GACACAACAG CTCACCCCTT TGCCGTGCGC CATCACCGTG
151 GACGACGACC TAGATGACCA TAGCCCAGAT GATGTGGATG CTCACGCCGG
201 TGCCTCCTCA CCCCTTCAAA TTTCTAGCTC TGATGAAGCT AGGGCTCCCA
251 CCAGACGCTC CGAGGTCAAG ATCGAAACTG ATTTAGTCTC CTCAGCCTGT
301 ACCCATTATG CCTGTGATGA CGTCCGTGGC AAGGGGAAGA ACAAAGGGAC
351 CACCAAGGAT GTTGGGAGGC TAAATAGGGT ATCAAAGGCC TCATCCTTTG
401 TTGATTCTTA TTCCGATTCT GATTATGACG ACTGCGAGGA GGACCAAGGA
451 ACAAGAACAG ATTATGCTGT TAAGCAGCTA AGAAGCAAGG GATTCACAAG
501 GAGACCACCC AACACCCCAA CATTCAGGAA CCATGGTGTC AGCGACGATG
551 AGCTGGGTCA GGAGAAGGAG AACCTTGGAG CTGTGGAGAA CAATGCTGAG
601 GATGTTGGAT GGGGAGAAGA CAGAGGACTT CAAGATGGAT CCAACTGGAA
651 CTGCTGCAAC ATCCAAGCCA TACAAGCTCC CAGGAAAGAT ATTCAAGATG
701 CTTTTCGCCC ACCAGCGCGA GGGCCTCCGA TGGCTCTGGG TTCTGCACTG
751 CAGGGGAACA GGAGGAATCC TAGGGGATGA CATGGGTCTT GGCAAGACGA
801 TGCAGGTTGC TGCATTTTTG GCTGGACTGT TTCATTCTCG TCTAGTCAAG
851 AGGGTGCTCA TTGTTGCTCC AAAGACACTT CTGGCCCATT GGACAAAGGA
901 GCTTTCAATT GTTGGCCTTA AAGAAAAGAT CAGAGACTAC TCTGGCCCCA
951 GCACAAATAT TCGCAATTAT GAACTCCAAT ATGCCTTCAA GGAGGGTGGT
1001 ATCCTCATAA CCACCTATGA CATTGTCAGG AACAACTACA AGCTCATAAG
1051 AGGCAACTCC TACAACAACA GCAATGATGA TGATGATGAG GAAGGAACTT
1101 TGTGGAATTA CGTAATTCTT GATGAGGGAC ATCTAATAAA AAATAATAAG
1151 ACACAAAGGG CCCAAAGTTT GTACGAAATA CCTTGTGCCC ATCGCATTGT
1201 GATCAGTGGA ACACCTATTC AAAATAACTT GAAGGAAATG TGGACTCTGT
1251 TCAATTTCTG TTGCCCAGAT GTCTTGGGTG ATAAACAGCA GTTCAAAATA
1301 AGGTATGAAA CGGCTATCCT TCGAGGAAAT GACAAAAATG CTACCGCTCG
1351 AGAGAAGCAC GTAGGCTCAA ATGTAGCAAA GGAACTAAGA GAGCGAATCA
1401 AGCCATACTT TTTGCGGCGC CTGAAAAGTG AAGTTGTCTT TGATACTGGT
1451 GCATCAGAAG AAAAAACATT AGCCAAGAAG AATGAGCTAA TTGTCTGGCT
1501 GAAGTTAACA CCATGCCAGA GGAAACTATA TGAAGCTTTT CTAAATAGTG
1551 AGCTGGTTCA TTTAGCATTG CAGCCAAAGG CATCACCGTT GGCTGCAATC
1601 ACAATATTGA AGAAAATATG TGATCATCCA CTGCTATTAA CTAAGAAAGG
1651 TGCTGAGGGT GTGTTGGAAG GAATGGGTGA AATGTTGAAT GATCAAGACA
1701 TTGGAATGGT GGAAAAAATG GCCATGAACC TTGCAGATAT GGCTCATGAT
1751 GATAATGCAC TGGAAGTTGG TCAGGATGTC TCATGCAAGC TATCATTCAT
1801 CATGTCCTTG TTGCGGAACC TTGTTGGAGA GGGGCATCAT GTTTTAATAT
1851 TTTCACAGAC TCGTAAAATG CTAAACCTTA TTCAGGAAGC TATAATATTA

1901 GAGGGCTATG CGTTTTTGCG CATTGATGGC ACCACCAAGG TTTCTGACCG
1951 GGAAAGGATT GTGAAGGACT TCCAAGAGGG TTGTGGAGCT CCAGTTTTTC
2001 TGCTAACCAC ACAAGTTGGT GGGCTTGGAC TTACACTCAC CAAGGCAACT
2051 CGTGTCATTG TAGTTGATCC TGCATGGAAC CCTAGTACAG ACAATCAAAG
2101 TGTTGATCGT GCTTACCGAA TTGGACAGAC TAAAAATGTG ATTGTATACC
2151 GCTTGATGAC ATCTGCGACC ATTGAAGAAA AGATATACAA ATTGCAGGTT
2201 TTGAAGGGCG CTCTGTTCAG GACAGCTACG GAGCAAAAAG AGCAAACACG
2251 TTACTTCAGC AAGAGTGAGA TTCAAGAGCT ATTTAGTTTG CCACAACAAG
2301 GATTTGATGT TTCCCTCACA CATAAGCAGT TGCAAGAAGA GCATGGTCAA
2351 CAAGTTGTTC TGGATGAGTC CTTGAGGAAG CATATACAGT TTCTGGAGCA
2401 ACAAGGAATA GCCGGTGTGA GTCATCACAG CCTCCTATTC TCTAAAACTG
2451 CAACCCTGCC CACTCTGAGT GAGAATGATG CACTGGACAG CAAACCTCGG
2501 GGCATGCCCA TGATGCCCCA GCAATATTAC AAGGGATCCT CATCTGACTA
2551 TGTCGCCAAC GGGGCATCTT TTGCGCTGAA GCCAAAGGAT GAAAGTTTCA
2601 CTGTTCGAAA CTACATTCCA AGTAACAGAA GCGCAGAGAG TCCTGAAGAG
2651 ATAAAGGCAA GAATCAACCG GCTTTCACAG ACCCTCTCCA ACGCTGTGCT
2701 GTTGTCGAAG CTACCAGATG GTGGTGAGAA GATAAGGAGG CAGATAAATG
2751 AGCTGGACGA GAAGCTGACT TCTGCTGAGA AGGGGCTGAA GGAGGGGGGC
2801 ACTGAAGTGA TTTCCTTGGA TGACTGATCC AAGACATGGA GAGTCTGTGC
2851 TCGGCAAAAG TAAA

Figure 5: AtIno80 and related sequences

Figure 5A

>AtIno80 coding sequence and derived protein

ATGGATCCTTCAAGACGACCACCGAAGGACTCTCCTTACGCGAATCTATTCGATCTCGA
GCCGTTGATGAAGTTTAGAATTCCGAAACCTGAAGATGAAGTTGATTATTATGGGAGTA
GTAGCCAGGATGAAAGTAGAAGCACTCaaggtggggtagtggcaaactacagcaatgggtctaaatcgaga
atgaatgcgagctccaagaagagaaagcggtggacagaagctgaggatgcagaggacgatgatgatctctacaatcaacat
gttactgaggagcactaccgatcaatgcttggggagcatgtacaaaaattcaaaaataggtccaaggagactcaagggaatcc
tcctcatctgatgggttttccggtgctaaagagcaatgtgggcagttacagaggtaggaaaccagggaatgattaccatgggag
gttctatgacatggacaactctccaaattttgcagctgatgtgaccccacataggcgaggaagctaccatgatcgtgatattacac
ccaagatagcatatgaaccttcgtatttggacattggtgatggtgtcatctacaaaatccccccaagttatgacaagctggtggcat
cattaaacttaccgagcttttcagacattcatgtggaagaattttacttgaaaggaactctggatctGAGATCATTAGCAGA
ACTGATGGCAAGTGATAAAAGGTCTGGAGTAAGAAGCCGTAATGGAATGGGTGAGCCT
CGACCTCAATATGAATCTCTTCAAGCTAGAATGAAGGCCCTGTCACCTTCAAACTCCAC
CCCAAATTTTAGCCTCAAGGTGTCAGAAGCTGCAATGAATTCTGCCATTCCAGAAGGAT
CTGCTGGAAGTACTGCACGGACAATTCTGTCTGAGGGTGGTGTTTTACAGGTCCATTAC
GTGAAGATTCTGGAGAAGGGGGATACATACGAGATTGTTAAACGAAGTCTACCGAAGA
AGCTGAAAGCAAAGAATGATCCTGCAGTCATTGAGAAAACAGAAAGGGATAAAATTAGA
AAAGCCTGGATCAATATTGTCAGAAGAGATATAGCAAAACACCATAGAATTTTCACTACT
TTTCATCGTAAACTATCAATTGATGCCAAGAGGTTTGCAGATGGTTGCCAAAGAGAGGT
GAGAATGAAGGTGGGTAGATCATACAAAATCCCAAGAACTGCACCAATTCGCACTAGGA
AGATATCCAGAGACATGCTGCTATTCTGGAAGCGATATGACAAGCAGATGGCAGAAGA
GAGGAAAAAGCAAGAAAAGGAAGCTGCAGAGGCTTTTAAACGTGAACAGGAGCAGCGA
GAGTCAAAAAGGCAGCAACAAAGGCTCAATTTCCTTATTAAACAGACTGAGCTTTACAG
TCACTTCATGCAAAACAAGACCGATTCGAATCCTTCCGAAGCCTTACCAATAGGTGATG
AAAATCCGATTGACGAAGTGCTCCCAGAAACTTCAGCGGCAGAACCTTCTGAGGTAGA
GGATCCTGAAGAGGCTGAACTGAAGGAAAAGGTCTTGAGAGCTGCCCAAGATGCGGTG
TCTAAGCAGAAGCAAATAACAGATGCATTTGACACTGAATATATGAAGCTACGCCAAACT

TCTGAAATGGAAGGTCCTTTAAATGATATATCAGTTTCTGGCTCGAGCAATATAGATTTG
CATAACCCATCTACAATGCCTGTTACATCAACAGTTCAGACTCCAGAGTTATTTAAAGGA
ACCCTTAAAGAATACCAAATGAAAGGCCTTCAGTGGCTAGTCAATTGTTATGAGCAGGG
TTTGAATGGCATACTTGCTGATGAAATGGGCTTGGGTAAGACTATTCAAGCTATGGCGT
TCTTGGCACATTTGGCTGAGGAAAAGAACATTTGGGGTCCATTTCTTGTTGTTGCCCCT
GCCTCTGTTCTTAACAATTGGGCTGATGAAATCAGTCGTTTCTGTCCTGACTTGAAAACT
CTTCCATATTGGGGAGGATTACAAGAACGAACAATTTTAAGAAAGAATAtcaatcccaagcgtat
gtaccgaagggatgctggctttcatattttgattactagctatcagctattagtcactgatgaaaagtattttcgccgggtgaagtggc
aatatatggtgctagatgaggcccaagcaatcaagagttcctccagtataagatggaaaaccttcttagttttaactgtcggaac
cgattgcttctgactggtactccaattcagaacaacatggcagagttatgggccctgctgcatttcatcatgccaatgttgtttgacaa
ccatgatcaatttaatgaatggttctcaaaaggaattgagaatcatgctgaacacggaggcactttaaatgagcaccagcttaac
agactgcatgcgatcttgaaaccgttcatgcttcgacgggtaaaaaaggatgtggtttctgagctaactacaaagacggaagtta
cagtacactgcaagctcagttctcgacaacaagctttttATCAGGCTATTAAGAACAAAATTTCTCTGGCTG
AGTTGTTTGATAGCAACCGCGGACAATTTACTGATAAGAAAGTATTGAATTTAATGAATA
TTGTCATTCAACTAAGGAAGGTTTGCAACCATCCAGAGTTGTTCGAAAGGAATGAAGGG
AGCTCGTATCTCTACTTTGGAGTGACTTCCAATTCTCTTTTGCCCCATCCCTTTGGTGAG
CTAGAGGATGTACATTATTCTGGTGGTCAAAATCCGATAATATACAAGATACCTAAGCTA
CTACACCAAGAGGTGCTCCAAAATTCTGAAACATTTTGTTCTTCTGTCGGGCGTGGCAT
CTCAAGAGAATCTTTTCTGAAGCATTTTAATATATATTCACCTGAGTATATTCTTAAGTCA
ATATTCCCATCTGATAGTGGGGTAGATCAAGTGGTTAGTGGAAGTGGAGCATTTGGCTT
TTCACGCTTGATGGATCTATCACCATCAGAAGTTGGATATCTGGCTCTGTGTTCTGTTG
CAGAAAGGCTATTATTTTCTATACTGAGGTGGGAGCGGCAATTTTTGGATGAATTAGTT
AACTCTCTTATGGAGTCCAAGGATGGTGATCTTAGTGACAATAACATCGAGAGAGTTAA
AACCAAAGCTGTCACAAGAATGTTGCTGATGCCATCAAAAGTTGAAACGAATTTTCAGAA
AAGGAGACTAAGCACAGGGCCTACCCGTCCTTCATTTGAAGCGCTAGTGATCTCTCATC
AGGATAGGTTTCTTTCAAGTATCAAACTCCTGCATTCTGCATATACTTATATCCCAAAAG
CCAGAGCTCCACCTGTAAGCATTCATTGCTCGGACAGAAATTCGGCATACAGAGTTACA
GAAGAATTACATCAACCATGGCTTAAGAGACTATTAATCGGTTTTGCACGAACGTCAGA
AGCTAATGGACCCAGGAAGCCTAACAGCTTTCCACATCCTTTAATCCAAGAAATTGATTC
AGAACTTCCAGTTGTGCAGCCTGCGCTTCAACTGACACACAGAATATTTGGTTCTTGCC
CTCCAATGCAAAGTTTTGACCCAGCAAAGTTGCTCACGGACTCTGGGAAGCTGCAGAC
ACTTGATATATTATTGAAGCGGCTTCGAGCTGGAAATCACAGGGTGCTCCTGTTTGCAC
AAATGACAAAGATGCTGAACATTCTCGAGGATTATATGAACTATAGAAAGTACAAGTACC
TCAGGCTTGATGGATCCTCCACCATCATGGATCGCCGAGATATGGTTAGGGATTTTCAG
CATAGGAGCGATATTTTTGTATTCTTGCTGAgcaccagagctggaggacttggtatcaacttgacggctgc
agacactgtcattttctatgaaagtgattggaatcccaccttggatttacaagctatggacagggctcatcgtcttggacagacaaa
agatgttactgtttatcgtctcatctgtaaggagacggtggaagagaaaattttgcacagggcaagtcagaaaaatacagttcaa
cagcttgttatgactggagggcatgttcagggtgatgattttcttggagctgcggatgtggtatctctgctaatggatgatgcggagg
cagcacaactggagcagaaattcagagaactaccattacaggtaaaggacaggcagaagaaaaagacgaAACGTAT
CAGAATAGATGCTGAAGGAGATGCAACTTTGGAAGAGTTAGAAGATGTTGACCGACAG
GATAACGGACAGGAACCTTTGGAAGAACCGGAAAAGCCAAAATCCAGTAATAAAAAGAG
GAGAGCTGCTTCAAATCCGAAAGCTAGAGCTCCTCAGAAAGCAAAGGAAGAAGCAAAT
GGTGAAGATACTCCTCAGAGGACAAAAAGGGTAAAGAGACAAACAAAGAGCATAAACG
AAAGTCTTGAACCTGTATTCTCTGCCTCTGTAACAGAATCAAATAAAGGATTCGATCCAA
GTAGCTCCGCTAACTAA

Figure 5B

>Derived AtIno80 protein sequence

MDPSRRPPKDSPYANLFDLEPLMKFRIPKPEDEVDYYGSSSQDESRSTQG
GVVANYSNGSKSRMNASSKKRKRWTEAEDAEDDDDLYNQHVTEEHYRSML
GEHVQKFKNRSKETQGNPPHLMGFPVLKSNVGSYRGRKPGNDYHGRFYDM

DNSPNFAADVTPHRRGSYHDRDITPKIAYEPSYLDIGDGVIYKIPPSYDK
LVASLNLPSFSDIHVEEFYLKGTLDLRSLAELMASDKRSGVRSRNGMGEP
RPQYESLQARMKALSPSNSTPNFSLKVSEAAMNSAIPEGSAGSTARTILS
EGGVLQVHYVKILEKGDTYEIVKRSLPKKLKAKNDPAVIEKTERDKIRKA
WINIVRRDIAKHHRIFTTFHRKLSIDAKRFADGCQREVRMKVGRSYKIPR
TAPIRTRKISRDMLLFWKRYDKQMAEERKKQEKEAAEAFKREQEQRESKR
QQQRLNFLIKQTELYSHFMQNKTDSNPSEALPIGDENPIDEVLPETSAAE
PSEVEDPEEAELKEKVLRAAQDAVSKQKQITDAFDTEYMKLRQTSEMEGP
LNDISVSGSSNIDLHNPSTMPVTSTVQTPELFKGTLKEYQMKGLQWLVNC
YEQGLNGILADEMGLGKTIQAMAFLAHLAEEKNIWGPFLVVAPASVLNNW
ADEISRFCPDLKTLPYWGGLQERTILRKNINPKRMYRRDAGFHILITSYQ
LLVTDEKYFRRVKWQYMVLDEAQAIKSSSSIRWKTLLSFNCRNRLLLTGT
PIQNNMAELWALLHFIMPMLFDNHDQFNEWFSKGIENHAEHGGTLNEHQL
NRLHAILKPFMLRRVKKDVVSELTTKTEVTVHCKLSSRQQAFYQAIKNKI
SLAELFDSNRGQFTDKKVLNLMNIVIQLRKVCNHPELFERNEGSSYLYFG
VTSNSLLPHPFGELEDVHYSGGQNPIIYKIPKLLHQEVLQNSETFCSSVG
RGISRESFLKHFNIYSPEYILKSIFPSDSGVDQVVSGSGAFGFSRLMDLS
PSEVGYLALCSVAERLLFSILRWERQFLDELVNSLMESKDGDLSDNNIER
VKTKAVTRMLLMPSKVETNFQKRRLSTGPTRPSFEALVISHQDRFLSSIK
LLHSAYTYIPKARAPPVSIHCSDRNSAYRVTEELHQPWLKRLLIGFARTS
EANGPRKPNSFPHPLIQEIDSELPVVQPALQLTHRIFGSCPPMQSFDPAK
LLTDSGKLQTLDILLKRLRAGNHRVLLFAQMTKMLNILEDYMNYRKYKYL
RLDGSSTIMDRRDMVRDFQHRSDIFVFLLSTRAGGLGINLTAADTVIFYE
SDWNPTLDLQAMDRAHRLGQTKDVTVYRLICKETVEEKILHRASQKNTVQ
QLVMTGGHVQGDDFLGAADVVSLLMDDAEAAQLEQKFRELPLQVKDRQKK
KTKRIRIDAEGDATLEELEDVDRQDNGQEPLEEPEKPKSSNKKRRAASNP
KARAPQKAKEEANGEDTPQRTKRVKRQTKSINESLEPVFSASVTESNKGF
DPSSSAN*

Figure 5C
>Alignment of AtIno 80 sequence and public sequence, At3g57300, showing splicing difference

Query: claimed sequence
Sbjct: gi|18410689|ref|NM_115590.1| (AGI:At3g57300)

```
Query: 1    atggatccttcaagacgaccaccgaaggactctccttacgcgaatctattcgatctcgag 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1    atggatccttcaagacgaccaccgaaggactctccttacgcgaatctattcgatctcgag 60

Query: 61   ccgttgatgaagtttagaattccgaaacctgaagatgaagttgattattatgggagtagt 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 61   ccgttgatgaagtttagaattccgaaacctgaagatgaagttgattattatgggagtagt 120

Query: 121  agccaggatgaaagtagaagcactcaaggtggggtagtggcaaactacagcaatgggtct 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 121  agccaggatgaaagtagaagcactcaaggtggggtagtggcaaactacagcaatgggtct 180

Query: 181  aaatcgagaatgaatgcgagctccaagaagagaaagcggtggacagaagctgaggatgca 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 181  aaatcgagaatgaatgcgagctccaagaagagaaagcggtggacagaagctgaggatgca 240

Query: 241  gaggacgatgatgatctctacaatcaacatgttactgaggagcactaccgatcaatgctt 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 241  gaggacgatgatgatctctacaatcaacatgttactgaggagcactaccgatcaatgctt 300

Query: 301  ggggagcatgtacaaaaattcaaaaataggtccaaggagactcaagggaatcctcctcat 360
```

```
Sbjct: 301  ggggagcatgtacaaaaattcaaaaataggtccaaggagactcaagggaatcctcctcat 360

Query: 361  ctgatgggttttccggtgctaaagagcaatgtgggcagttacagaggtaggaaaccaggg 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 361  ctgatgggttttccggtgctaaagagcaatgtgggcagttacagaggtaggaaaccaggg 420

Query: 421  aatgattaccatgggaggttctatgacatggacaactctccaaattttgcagctgatgtg 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 421  aatgattaccatgggaggttctatgacatggacaactctccaaattttgcagctgatgtg 480

Query: 481  accccacataggcgaggaagctaccatgatcgtgatattacacccaagatagcatatgaa 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 481  accccacataggcgaggaagctaccatgatcgtgatattacacccaagatagcatatgaa 540

Query: 541  ccttcgtatttggacattggtgatggtgtcatctacaaaatccccccaagttatgacaag 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 541  ccttcgtatttggacattggtgatggtgtcatctacaaaatccccccaagttatgacaag 600

Query: 601  ctggtggcatcattaaacttaccgagcttttcagacattcatgtggaagaattttacttg 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 601  ctggtggcatcattaaacttaccgagcttttcagacattcatgtggaagaattttacttg 660

Query: 661  aaaggaactctggatctgagatcattagcagaactgatggcaagtgataaaaggtctgga 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 661  aaaggaactctggatctgagatcattagcagaactgatggcaagtgataaaaggtctgga 720

Query: 721  gtaagaagccgtaatggaatgggtgagcctcgacctcaatatgaatctcttcaagctaga 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 721  gtaagaagccgtaatggaatgggtgagcctcgacctcaatatgaatctcttcaagctaga 780

Query: 781  atgaaggccctgtcaccttcaaactccaccccaaattttagcctcaaggtgtcagaagct 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 781  atgaaggccctgtcaccttcaaactccaccccaaattttagcctcaaggtgtcagaagct 840

Query: 841  gcaatgaattctgccattccagaaggatctgctggaagtactgcacggacaattctgtct 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 841  gcaatgaattctgccattccagaaggatctgctggaagtactgcacggacaattctgtct 900

Query: 901  gagggtggtgttttacaggtccattacgtgaagattctggagaaggggggatacatacgag 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 901  gagggtggtgttttacaggtccattacgtgaagattctggagaaggggggatacatacgag 960

Query: 961  attgttaaacgaagtctaccgaagaagctgaaagcaaagaatgatcctgcagtcattgag 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 961  attgttaaacgaagtctaccgaagaagctgaaagcaaagaatgatcctgcagtcattgag 1020

Query: 1021 aaaacagaaagggataaaattagaaaagcctggatcaatattgtcagaagagatatagca 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1021 aaaacagaaagggataaaattagaaaagcctggatcaatattgtcagaagagatatagca 1080

Query: 1081 aaacaccatagaattttcactacttttcatcgtaaactatcaattgatgccaagaggttt 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1081 aaacaccatagaattttcactacttttcatcgtaaactatcaattgatgccaagaggttt 1140

Query: 1141 gcagatggttgccaaagagaggtgagaatgaaggtgggtagatcatacaaaatcccaaga 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1141 gcagatggttgccaaagagaggtgagaatgaaggtgggtagatcatacaaaatcccaaga 1200

Query: 1201 actgcaccaattcgcactaggaagatatccagagacatgctgctattctggaagcgatat 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1201 actgcaccaattcgcactaggaagatatccagagacatgctgctattctggaagcgatat 1260

Query: 1261 gacaagcagatggcagaagagaggaaaaagcaagaaaaggaagctgcagaggcttttaaa 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1261 gacaagcagatggcagaagagaggaaaaagcaagaaaaggaagctgcagaggcttttaaa 1320

Query: 1321 cgtgaacaggagcagcgagagtcaaaaaggcagcaacaaaggctcaatttccttattaaa 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1321 cgtgaacaggagcagcgagagtcaaaaaggcagcaacaaaggctcaatttccttattaaa 1380

Query: 1381 cagactgagctttacagtcacttcatgcaaaacaagaccgattcgaatccttccgaagcc 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1381 cagactgagctttacagtcacttcatgcaaaacaagaccgattcgaatccttccgaagcc 1440

Query: 1441 ttaccaataggtgatgaaaatccgattgacgaagtgctcccagaaacttcagcggcagaa 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct: 1441 ttaccaataggtgatgaaaatccgattgacgaagtgctcccagaaacttcagcgg agaa 1500

Query: 1501 ccttctgaggtagaggatcctgaagaggctgaactgaaggaaaaggtcttgaga gctgcc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1501 ccttctgaggtagaggatcctgaagaggctgaactgaaggaaaaggtcttgaga gctgcc 1560

Query: 1561 caagatgcggtgtctaagcagaagcaaataacagatgcatttgacactgaatata gaag 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1561 caagatgcggtgtctaagcagaagcaaataacagatgcatttgacactgaatata gaag 1620

Query: 1621 ctacgccaaacttctgaaatggaaggtcctttaaatgatatatcagtttctggct gagc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1621 ctacgccaaacttctgaaatggaaggtcctttaaatgatatatcagtttctggct gagc 1680

Query: 1681 aatatagatttgcataacccatctacaatgcctgttacatcaacagttcagactc agag 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1681 aatatagatttgcataacccatctacaatgcctgttacatcaacagttcagactc agag 1740

Query: 1741 ttatttaaaggaacccttaaagaataccaaatgaaaggccttcagtggctagtca attgt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1741 ttatttaaaggaacccttaaagaataccaaatgaaaggccttcagtggctagtca attgt 1800

Query: 1801 tatgagcagggtttgaatggcatacttgctgatgaaatgggcttgggtaagacta ttcaa 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1801 tatgagcagggtttgaatggcatacttgctgatgaaatgggcttgggtaagacta ttcaa 1860

Query: 1861 gctatggcgttcttggcacatttggctgaggaaaagaacatttggggtccatttc ttgtt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1861 gctatggcgttcttggcacatttggctgaggaaaagaacatttggggtccatttc ttgtt 1920

Query: 1921 gttgcccctgcctctgttcttaacaattgggctgatgaaatcagtcgtttctgtc tgac 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1921 gttgcccctgcctctgttcttaacaattgggctgatgaaatcagtcgtttctgtc tgac 1980

Query: 1981 ttgaaaactcttccatattggggaggattacaagaacgaacaattttaagaaa gaatatc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1981 ttgaaaactcttccatattggggaggattacaagaacgaacaattttaagaaa gaatatc 2040

Query: 2041 aatcccaagcgtatgtaccgaagggatgctggctttcatattttgattactagctatcag 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2041 aatcccaagcgtatgtaccgaagggatgctggctttcatattttgattactagctatcag 2100

Query: 2101 ctattagtcactgatgaaaagtattttcgccgggtgaagtggcaatatatggtgctagat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2101 ctattagtcactgatgaaaagtattttcgccgggtgaagtggcaatatatggtgctagat 2160

Query: 2161 gaggcccaagcaatcaagagttcctccagtataagatggaaaaccccttcttagttttaac 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2161 gaggcccaagcaatcaagagttcctccagtataagatggaaaaccccttcttagttttaac 2220

Query: 2221 tgtcggaaccgattgcttctgactggtactccaattcagaacaacatggcagagttatgg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2221 tgtcggaaccgattgcttctgactggtactccaattcagaacaacatggcagagttatgg 2280

Query: 2281 gccctgctgcatttcatcatgccaatgttgtttgacaaccatgatcaatttaatgaatgg 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2281 gccctgctgcatttcatcatgccaatgttgtttgacaaccatgatcaatttaatgaatgg 2340

Query: 2341 ttctcaaaaggaattgagaatcatgctgaacacggaggcactttaaatgagcaccagctt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2341 ttctcaaaaggaattgagaatcatgctgaacacggaggcactttaaatgagcaccagctt 2400

Query: 2401 aacagactgcatgcgatcttgaaaccgttcatgcttcgacgggtaaaaaaggatgtggtt 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2401 aacagactgcatgcgatcttgaaaccgttcatgcttcgacgggtaaaaaaggatgtggtt 2460

Query: 2461 tctgagctaactacaaagacggaagttacagtacactgcaagctcagttctcgacaacaa 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2461 tctgagctaactacaaagacggaagttacagtacactgcaagctcagttctcgacaacaa 2520

Query: 2521 gctttttatcaggctattaagaacaaaatttctctggctgagttgtttgatagcaaccgc 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2521 gctttttatcaggctattaagaacaaaatttctctggctgagttgtttgatagcaaccgc 2580

Query: 2581 ggacaatttactgataagaaagtattgaatttaatgaatattgtcattcaactaaggaag 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2581 ggacaatttactgataagaaagtattgaatttaatgaatattgtcattcaactaaggaag 2640
```

```
Query: 2641 gtttgcaaccatccagagttgttcgaaaggaatgaagggagctcgtatctctactttgga 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2641 gtttgcaaccatccagagttgttcgaaaggaatgaagggagctcgtatctctactttgga 2700

Query: 2701 gtgacttccaattctcttttgccccatccctttggtgagctagaggatgtacattattct 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2701 gtgacttccaattctcttttgccccatccctttggtgagctagaggatgtacattattct 2760

Query: 2761 ggtggtcaaaatccgataatatacaagatacctaagctactacaccaagaggtgctccaa 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2761 ggtggtcaaaatccgataatatacaagatacctaagctactacaccaagaggtgctccaa 2820

Query: 2821 aattctgaaacattttgttcttctgtcgggcgtggcatctcaagagaatcttttctgaag 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2821 aattctgaaacattttgttcttctgtcgggcgtggcatctcaagagaatcttttctgaag 2880

Query: 2881 cattttaatatatattcacctgagtatattcttaagtcaatattcccatctgatagtggg 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2881 cattttaatatatattcacctgagtatattcttaagtcaatattcccatctgatagtggg 2940

Query: 2941 gtagatcaagtggttagtggaagtggagcatttggcttttcacgcttgatggatctatca 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2941 gtagatcaagtggttagtggaagtggagcatttggcttttcacgcttgatggatctatca 3000

Query: 3001 ccatcagaagttggatatctggctctgtgttctgttgcagaaaggctattattttctata 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3001 ccatcagaagttggatatctggctctgtgttctgttgcagaaaggctattattttctata 3060

Query: 3061 ctgaggtgggagcggcaatttttggatgaattagttaactctcttatggagtccaaggat 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3061 ctgaggtgggagcggcaatttttggatgaattagttaactctcttatggagtccaaggat 3120

Query: 3121 ggtgatcttagtgacaataacatcgagagagttaaaaccaaagctgtcacaagaatgttg 3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3121 ggtgatcttagtgacaataacatcgagagagttaaaaccaaagctgtcacaagaatgttg 3180

Query: 3181 ctgatgccatcaaaagttgaaacgaattttcagaaaaggagactaagcacagggcctacc 3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3181 ctgatgccatcaaaagttgaaacgaattttcagaaaaggagactaagcacagggcctacc 3240

Query: 3241 cgtccttcatttgaagcgctagtgatctctcatcaggataggtttctttcaagtatcaaa 3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3241 cgtccttcatttgaagcgctagtgatctctcatcaggataggtttctttcaagtatcaaa 3300

Query: 3301 ctcctgcattctgcatatacttatatcccaaaagccagagctccacctgtaagcattcat 3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3301 ctcctgcattctgcatatacttatatcccaaaagccagagctccacctgtaagcattcat 3360

Query: 3361 tgctcggacagaaattcggcatacagagttacagaagaattacatcaaccatggcttaag 3420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3361 tgctcggacagaaattcggcatacagagttacagaagaattacatcaaccatggcttaag 3420

Query: 3421 agactattaatcggttttgcacgaacgtcagaagctaatggacccaggaagcctaacagc 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3421 agactattaatcggttttgcacgaacgtcagaagctaatggacccaggaagcctaacagc 3480

Query: 3481 tttccacatcctttaatccaagaaattgattcagaacttccagttgtgcagcctgcgctt 3540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3481 tttccacatcctttaatccaagaaattgattcagaacttccagttgtgcagcctgcgctt 3540

Query: 3541 caactgacacacagaatatttggttcttgccctccaatgcaaagttttgacccagcaaag 3600
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3541 caactgacacacagaatatttggttcttgccctccaatgcaaagttttgacccagcaaag 3600

Query: 3601 ttgctcacggactctgggaagctgcagacacttgatatattattgaagcggcttcgagct 3660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3601 ttgctcacggactctgggaagctgcagacacttgatatattattgaagcggcttcgagct 3660

Query: 3661 ggaaatcacagggtgctcctgtttgcacaaatgacaaagatgctgaacattctcgaggat 3720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3661 ggaaatcacagggtgctcctgtttgcacaaatgacaaagatgctgaacattctcgaggat 3720

Query: 3721 tatatgaactatagaaagtacaagtacctcaggcttgatggatcctccaccatcatggat 3780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3721 tatatgaactatagaaagtacaagtacctcaggcttgatggatcctccaccatcatggat 3780

Query: 3781 cgccgagatatggttagggattttcagcataggagcgatatttttgtattcttgctgagc 3840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3781 cgccgagatatggttagggattttcagcataggagcgatatttttgtattcttgctgagc 3840

Query: 3841 accagagctggaggacttggtatcaacttgacggctgcagacactgtcattttctatgaa 3900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3841 accagagctggaggacttggtatcaacttgacggctgcagacactgtcattttctatgaa 3900

Query: 3901 agtgattggaatcccaccttggatttacaagctatggacagggctcatcgtcttggacag 3960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3901 agtgattggaatcccaccttggatttacaagctatggacagggctcatcgtcttggacag 3960

Query: 3961 acaaaagatg 3970
            ||||||||||
Sbjct: 3961 acaaaagatg 3970


 Score = 1001 bits (505), Expect = 0.0
 Identities = 522/528 (98%), Gaps = 6/528 (1%)
 Strand = Plus / Plus

Query: 3997 gagacggtggaagagaaaattttgcacagggcaagtcagaaaaatacagttcaacagctt 4056
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3970 gagacggtggaagagaaaattttgcacagggcaagtcagaaaaatacagttcaacagctt 4029

Query: 4057 gttatgactggagggcatgttcagggtgatgattttcttggagctgcggatgtggtatct 4116
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4030 gttatgactggagggcatgttcagggtgatgattttcttggagctgcggatgtggtatct 4089

Query: 4117 ctgctaatggatgatgcggaggcagcacaactggagcagaaattcagagaactaccatta 4176
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4090 ctgctaatggatgatgcggaggcagcacaactggagcagaaattcagagaactaccatta 4149

Query: 4177 caggtaaaggacaggcagaagaaaaagacgaaacgtatcagaatagatgctgaaggagat 4236
            ||||      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4150 cagg------acaggcagaagaaaaagacgaaacgtatcagaatagatgctgaaggagat 4203

Query: 4237 gcaactttggaagagttagaagatgttgaccgacaggataacggacaggaacctttggaa 4296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4204 gcaactttggaagagttagaagatgttgaccgacaggataacggacaggaacctttggaa 4263

Query: 4297 gaaccggaaaagccaaaatccagtaataaaaagaggagagctgcttcaaatccgaaagct 4356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4264 gaaccggaaaagccaaaatccagtaataaaaagaggagagctgcttcaaatccgaaagct 4323

Query: 4357 agagctcctcagaaagcaaaggaagaagcaaatggtgaagatactcctcagaggacaaaa 4416
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct: 4324 agagctcctcagaaagcaaaggaagaagcaaatggtgaagatactcctcagaggacaaaa 4383

Query: 4417 agggtaaagagacaaacaaagagcataaacgaaagtcttgaacctgtattctctgcctct 4476
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4384 agggtaaagagacaaacaaagagcataaacgaaagtcttgaacctgtattctctgcctct 4443

Query: 4477 gtaacagaatcaaataaaggattcgatccaagtagctccgctaactaa 4524
            ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4444 gtaacagaatcaaataaaggattcgatccaagtagctccgctaactaa 4491
```

Figure 6:

>AtRvb1 (At5g22330)

>2564051 CDS from MWD9 (protein BAB08331)

atggagaaagtaaagattgaagaaattcagtccaccgctaagaaacaacg
gattgctactcacacccatatcaaaggccttggcctcgagccaactggta
tccctataaaattggcagctggatttgttggtcaacttgaggctagagag
gcagctggtcttgtagttgacatgattaagcagaagaaaatggcgggcaa
ggctcttttgcttgctggacctcctggaactgggaaaacagctttggctc
ttggaatctctcaagagctgggaagcaaggttccattctgtccaatggtt
ggatctgaggtttactcatcagaggttaagaaaacagaggttctcatgga
gaattttagacgtgccattggtctacgtatcaaggaaaccaaagaagtct
atgaaggggaggtcaccgagctgtcaccagaagaaactgaaagcctcact
ggaggttatggtaaaagcatcagccatgttgtaattacactcaagacagt
caaaggaaccaaacatctgaaattggatcccactatctatgatgccttga
ttaaggaaaaggtagctgtaggagatgtaatctatatcgaagcaaacagt
ggagctgtcaaacgggtaggtagaagtgatgcttttgccactgaatttga
tctggaagcagaagaatatgttccacttcccaaaggagaggtccacaaaa
agaaagagatagtgcaggatgtcacactccaagatctggatgcagcaaat
gctcgacctcaaggtggccaggatatactttctttgatgggccaaatgat
gaaaccgcggaagactgagatcactgataagcttcggcaagaaattaaca
aggttgtgaaccgatatatagatgaaggtgtggcagagcttgttccagga
gttctatttattgatgaggttcatatgcttgatatggagtgcttctcata
cttgaaccgtgctcttgagagctcattatctccgatagtgatatttgcaa
caaatagaggtgtttgcaacgtaagagggactgatatgcccagcccccat
ggagtccctattgatctattagatcgattggttatcatccggactcaaat
ctatgatccctctgaaatgatccagattatagccattcgtgcgcaagttg
aagaattaaccgtggatgaagaatgcttggttctacttggggagattggg
caaagaacttcactaaggcacgctgtgcagcttctgtctcctgccagcat

tgtagcgaaaatgaatggccgtgacaatatttgcaaggctgatatagagg
aagtaacatcactctacttggatgctaaatcttcagcaaagcttttgcat
gagcaacaagaaaaatacatctcatga

Figure 7:

>AtRvb21 (At5g67630)

>At5g67630 and 3'UTR (prot BAB08471.1)

atggcggaactaaagctatcagagagtcgggacttaaccagagtcgagcg
aatcggcgcacactcacacatcagaggactaggtctcgactctgccctcg
agccgcgagctgtttccgaaggtatggtcggtcaagtgaaggcgcgtaaa
gccgccggtgtaatccttcagatgattagagaagggaaaatcgcgggtcg
ggctattctaatagcgggtcaacccggaacgggtaagacagcgattgcaa
tgggtatggcgaaatctcttggcttggaaactccttttgcgatgattgca
ggaagtgaaattttctcattagagatgtcaaagacagaagctttgactca
gtcttttcgtaaagcgattggtgttaggatcaaagaagagacagaggtta
ttgaaggagaagttgttgaggttcagattgataggcctgcttcttctggt
gttgcttccaagtcagggaagatgactatgaaaacgactgatatggaaac
tgtgtatgatatgggagctaagatgattgaggctttgaacaaggagaaag
tgcagagtggtgatgttattgccattgataaagctactgggaagattact
aagcttggaagatcgttttcgaggtctcgtgattatgatgctatgggtgc
gcagaccaagtttgtgcagtgccctgaaggtgagttgcagaagaggaaag
aggttgtacattgtgtcactcttcacgagattgatgttatcaacagcagg
acacaagggtttcttgccctttcactggcgatactggagaaatccgatc
agaagtccgggaacaaattgatacaaaagtagctgaatggagagaagaag
gaaaagcagagatagttcccggagttctcttcattgatgaagtccacatg
ctcgacatcgaatgcttctcattccttaaccgagctctagaaaacgaaat
gtcaccaatccttgtggtggcaacaaaccgaggagtgacgacaatccgtg
gcacaaaccagaaatcaccacacgggatcccgattgatctccttgaccgt
cttctcatcatcactacccaaccttacacagacgatgacataaggaagat
attagaaatccgttgccaagaggaagacgttgagatgaacgaagaggcca
aacagcttttgacattgatcggacgtgatacatctctaaggtatgcgatt
catcttataaccgcagctgcattgtcatgccagaaacggaaagggaaagt
cgtggaggttgaggatattcagagagtttacagactgttcttggatgtga
ggagatcgatgcagtatcttgttgagtatcagagtcagtatatgttcagt

gaaccaatcaaaaacgatgaagctgctgcagaagacgaacaagatgctat
gcagatctgaGGATCCACCTCTGTTTGCCTTATTTATCATGTTTCGTGGT
GATATGTATGATTAGGATGTTGAACTCGGATTTATGTTTTTTTTTTTTTA
AGTTGTGACGAGATTCGGTTCTAGAAAATGATTTAACCAAGTTCAATACA
GATCGGTTTGGTACAAAACAAAAAAAAAAAAAA

Figure.8:

>AtRvb22 (At3g49830)
>At3g49830 prediction (protein CAB66921.1)
atggcagaactaaggttatcagaaactcgagacttaactaggatcgaaag
aatcggagcacactcacacatacgaggtttaggtctcgactcagtactcg
agccacgagccgtatccgaaggaatggttggtcaaatcaaagcacgtaaa
gccgccggagtaaccctcgagttgatcagagacggcaaaatctcgggtcg
ggctatacttatagcgggtcaacccggaacgggtaaaatcgcaatagcaa
tgggtatagcaaaatcacttggacaagaaacaccattcactatgattgca
ggaagtgagatcttttctttagagatgtcaaagactgaagctttaactca
agcttttcgtaaagctattggtgttaggatcaaagaagagactgacgtga
tagaaggagaagttgtgacgatttcgattgatagacctgcttcttctggt
ggttctgtgaagaagactgggaagataacaatgaagacgactgatatgga
atctaattttgatttgggatggaaattgattgagccattggataaggaga
aagtacagagtggtgatgttattgttttggataggttttgtgggaagatt
actaagcttggaagatcttttacgaggtctagagattttgatgttatggg
ttcaaagactaagtttgtgcagtgccctgaaggtgagcttgagaagagga
aggaggttttgcattctgtcacacttcatgagattgatgttattaatagc
aggactcaagggtatctagccctcttcacaggtgatacaggcgagattcg
ttcagaaacccgagagcaaagcgatactaaagtggcagagtggagagaag
aagggaaagctgaaatagtacctggtgttctcttcattgatgaagtccat
atgcttgatatcgaatgcttctctttcctgaatagagctctcgaaaacga
tatgtcaccaatcctggtcgtggctacaaacagaggaatgacaacaatcc
gaggaacaaaccagatatcagcacatgggatcccaatcgattttcttgac
cgtcttcttattatcacaacacagccttacacacaagacgagatcagaaa
tattttagagatccgttgccaagaagaggatgtggagatgaacgaggaag
cgaaacagcttctgactttgatcggatgtaatacctcgcttaggtacgcg
attcatctaatcaatgcagctgccctagcttgcctgaaacgtaaagggaa

agtcgtagagattcaggacattgagagagtttatagattgtttttagaca
ccaagagatcgatgcagtacttggttgagcatgagagcgagtacttgttt
agcgtgcctataaaaaacacacaggaggctactgcaggagaagaaacaga
acacgaggccatggaagtttga

Figure 9: At3g57290

>eIF3e Ath mRNA AF285832 (protein AAG53613.1)
ccacgcgtccgtaagaagattttgccagtgcgggaagcggcggagattga
gagattagcgacgatggaggaaagcaaacagaactatgacctgacgccac
taatagcgcctaacctggacagacacttggtgtttcctatattcgagttc
cttcaagagcgtcagctttaccctgatgagcagatcctgaagtctaaaat
ccagcttttgaaccagacgaacatggttgattacgccatggatattcaca
agagtctctaccacactgaagacgctcctcaagaaatggtggagagaaga
acagaggttgtcgctaggctcaaatctttggaggaggctgctgcaccact
cgtgtcttttctttttgaaccctaacgctgtgcaggagctaagagctgaca
agcagtacaatctccaaatgctcaaggaacgctaccagattggtccagac
cagattgaggctttgtaccagtacgccaagtttcagtttgaatgtggcaa
ctattctggtgctgctgattatctttaccagtacaggaccctgtgctcta
accttgagaggagtttgagtgccttgtggggaaagctcgcatctgaaata
ttgatgcaaaactgggatattgctcttgaagagcttaaccgtctcaaaga
gattattgactcaaagagtttttcatcgccgttaaaccaggtgcagaaca
ggatttggttgatgcattggggtctgtatatctttttttaaccatgataat
ggaaggacacagatcattgatctttttaaccaagacaagtatctgaatgc
catccaaactagtgctccacacttgctgcgctacttggcaactgctttca
ttgtcaacaaaaggagaagaccacaattgaaagaattcattaaggtcatt
cagcaagagcactactcctacaaagatccaattatcgagttcctggcatg
tgtgtttgtcaattatgactttgatggggctcaaaagaagatgaaagagt
gtgaagaggtcattgtgaatgatccattccttggcaagcgagttgaggat
ggaaacttttcaactgtaccactgagagatgaatttcttgaaaatgcccg
cctattcgtctttgaaacctattgcaaaattcatcaaaggattgacatgg
gggtacttgctgaaaaattgaatctgaactatgaggaggccgagagatgg
attgtgaacctaatccgcacctcaaagcttgatgccaagattgattctga
gtcaggaactgtaatcatggagcctactcagcccaacgtgcatgagcagt
tgataaaccacaccaaaggcttatcaggacgaacatacaagttagtgaat

cagctcttggaacacacacaggcgcaagcaactcgctagtcaaaattttg
ctgtggaagcctttccttgataaaactcaccttcggttgactggaattat
tttctttttcttgctctgagttcacctttactttgaaaaagattattat
ggagttgttctatttgaaatgttggatccacagattggaacattttccaa
ccaaatcagcattttgtaaaaaaaaaaaaaaaaaaaaaaa

Figure 10: plant homologs of Hw17